

TECHNICAL UNIVERSITY OF KOŠICE
Faculty of Electrical Engineering and Informatics



Electrical Engineering and Inormatics III

Proceeding
of
the Faculty of Electrical Engineering and Informatics
of
the Technical University of Košice

Editor: Dušan Kocur

September, 2012

Košice, Slovak Republic

ISBN 978-80-553-0890-6

Electrical Engineering and Informatics III

Proceeding of the Faculty of Electrical Engineering and Informatics of the Technical University of Košice

Košice, Slovak Republic September, 2012 ISBN 978-80-553-0890-6

Contents

Weather Prediction Using Neural Networks
Ján ADAMČÁK, Rudolf JAKŠA 1

Riadenie robota rozpoznávaním ľudských pohybov pomocou zariadenia Kinect
Peter ADAMONDY, Peter PAPCUN, Ján JADLOVSKÝ 5

Využitie internetového marketingu pre zvýšenie predajnosti internetového obchodu
Martin ANDERKO, František BABIČ 10

Analýza informačného systému firmy na základe IT Scorecard
Mária BACHLEDOVÁ, Adela TUŠANOVÁ, Ján PARALIČ 16

Využitie umelej inteligencie v spracovaní obrazu
Matúš BAČO 22

Manažment sémantických znalostí
Lucia BRÉDOVÁ, Marián MACH 30

Návrh využitia metodológie METHONTOLOGY pre aplikáciu vybraného konceptuálneho modelu
Peter BUTKA 34

Porovnanie metodológií pre budovanie ontologických modelov
Peter BUTKA 40

Experimentálna identifikácia nelineárneho dynamického systému pomocou IDENT Tool v prostredí Matlab
Jakub ČERKALA, Anna JADLOVSKÁ 46

Ekonomické aspekty identifikácie malého prúdového motora
Marek ČERNEJ, Ladislav MADARÁSZ, Vladimír GAŠPAR 52

Analýza dosiahnuteľnosti stavu výrobných liniek použitím Petriho sieti
Matej ČOPÍK 59

Detekcia tvárí s využitím neurónových sietí
Dušan DOVALA, Rudolf JAKŠA 64

Towards the Optimization of IT Service Delivery Processes in Governmental Environment
Jana DUCHOVÁ, Karol FURDÍK 69

Riadenie robota pomocou platformy Android
Tomáš DUPLINSKÝ, Peter PAPCUN, Ján JADLOVSKÝ 75

Predikcia bankrotu firiem pomocou vhodných metód objavovania znalostí
Tomáš GROSS, František BABIČ 80

Využitie BI na analýzu finančnej situácie firiem
Cecília HAVRILOVÁ, František BABIČ 85

Využitie neurónovej siete FALCON pre riadenie pohybu robota Nao pri kope
Dušan HODOŠKO, Ján VAŠČÁK 90

Optimalizácia procesov v stavebnej firme
Lenka ILAVSKÁ, František BABIČ 101

Návrh a realizácia riadenia flexibilného montážneho podniku PLC automatom
Igor JADLOVSKÝ, Matej ČOPÍK, Ján SARNOVSKÝ 107

Adaptívne prediktívne riadenie na báze analytickej a experimentálnej identifikácie riadeného systému
Štefan JAJČIŠIN, Anna JADLOVSKÁ 111

Efektívne využívanie zdrojov vo výrobe pomocou procesného modelovania
Slavomíra JURČÍKOVÁ, František BABIČ 118

Automatická tvorba korpusov pre potreby analýzy sentimentu
Peter KONCZ, Ján PARALIČ 122

Znalostný sklad pre predmet objavovanie znalostí
Peter KONCZ, Ján PARALIČ 126

Prístupy k vnímaniu emócií v humanoidnej robotike
Peter KONCZ 129

Infuence of Learning Parameter γ on Stochastic Update Modifcations of Backpropagation Through Time
Juraj KOŠČÁK, Rudolf JAKŠA, Peter SINČÁK 135

Návrh a realizácia informačného systém pre flexibilný montážny podnik
Tomáš KRAJČI, Matej ČOPÍK, Ján JADLOVSKÝ 141

Riešenie úloh s ohraničeniami ako problém splniteľnosti
Michal LECHMAN, Marián MACH 145

Návrh a realizácia manažérskeho systému pre flexibilný montážny podnik
Pavel LINKESCH, Matej ČOPÍK, Ján JADLOVSKÝ 149

Fuzzy Cognitive Maps in Navigation of Robot Nao
Daniel LORENČÍK, Ján VAŠČÁK 154

Robustnosť komplexných sietí a ich tolerancia k poruchám a útokom
Gabriel LUKAČ, Marián MACH 161

Aplikácia dolovania dát vo výrobnom priemysle
Alexandra LUKÁČOVÁ 165

Social Web Mapping
Kristína MACHOVÁ, Tomáš RAKUŠČINEC 169

Návrh a realizácia riadenia križovatky PLC automatom
Gergő MARCI, Matej ČOPÍK, Ján SARNOVSKÝ 173

Využitie vhodných webových technológií pre prezentáciu a rezerváciu ubytovacích kapacít
Tomáš ORSÁG, František BABIČ 176

ISBN 978-80-553-0890-6 © 2012 FEI TUKE

Informačný systém pre online prieskumy spokojnosti
Beáta PETRÍKOVÁ, Iveta ZOLOTOVÁ 180

Použitie prostriedkov umelej inteligencie pri riešení stability pohybu robota Nao
Marián POLEK, Ján VAŠČÁK 185

Optimalizácia výrobného procesu prostredníctvom štíhlej výroby
Lukáš POPIK, František BABIČ 192

Návrh a realizácia aplikácie na analýzu výrobných procesov pomocou Petriho sietí
Rudolf REPČÍN, Matej ČOPÍK, Ján JADLOVSKÝ 196

Vyhľadávanie dokumentov s použitím formálnej konceptuálnej analýzy
Martin SARNOVSKÝ, Marek BALUCHA, Peter BUTKA 202

Vyhľadávanie webových stránok pomocou ekvivalencie tried v oblasti rough sets
Martin SARNOVSKÝ, Maroš MARCINOV, Peter BUTKA 208

Jednoduchý príklad modelovania autonómneho agenta v simulačnom prostredí Netlogo
Martin SARNOVSKÝ, Patrik VINCEJ, Peter BUTKA 214

Implementácia monitorovacieho nástroja pre cloudovú infraštruktúru
Martin SARNOVSKÝ, Peter SASARÁK 220

Luwak: Systém na extrakciu informácii z homogénnych textových dokumentov
Peter SMATANA 224

Comparison of MF ARTMAP Neural Network and MF ARTMAP Ceural Network with Modification of Fuzzy Relation
Peter SMOLÁR, Peter SINČÁK 228

Methods for the Number Modification of Clusters in Modified MF ARTMAP Neural Network
Peter SMOLÁR, Peter SINČÁK 234

Opinion Classification Based on Dictionary Generation
Peter SZABÓ, Kristína MACHOVÁ 240

Prečo monitorovať on-line komunikáciu?
Ján ŠTOFA, Kristína MACHOVÁ 244

Návrh a simulácia riadenia pre výukový model Magnetická levitácia
Peter ŠUSTER, Martin BENETIN, Anna JADLOVSKÁ 250

Možnosti nasadenia SaaS v konkrétnej firme
Daniela TOMKOVÁ, Adela TUŠANOVÁ, Ján PARALIČ 257

Optimalizácia výroby pomocou procesného modelovania
Veronika VARGOVÁ, František BABIČ 263

Evolučné prístupy v navigácii mobilných robotov
Ján VAŠČÁK 268

Use of Multi-agent Approaches in Robotic Soccer
Ján VAŠČÁK 277

ISBN 978-80-553-0890-6 © 2012 FEI TUKE

Návrh riadenia činnosti práčky pomocou fuzzy kognitívnych máp
Ján VAŠČÁK 284

Testovanie architektúr 1-Wire senzorových sietí
Jozef ADAMIŠIN, Ľudmila MACEKOVÁ 293

Základné princípy MIMO systému pre komunikáciu v automobilovej doprave
Dušan DEMIAN, Iveta GLADIŠOVÁ 299

Modified Simmons' Authentication Scheme Using Elliptic Curve Arithmetic and Combinatorial Design
Patrik GALLO, Dušan LEVICKÝ 305

DCT vizuálneho objektu pomocou ortogonalizácie jej bázových funkcií v oblastiach hraničných blokov
Iveta GLADIŠOVÁ, Ján MIHALÍK 310

Implementácia algoritmu mriežkového vektorového kvantovania
Iveta GLADIŠOVÁ 315

Effective Measurement Stand for Teaching Electronic Circuits
Marek GODLA, Linus MICHAELI, Ján ŠALIGA, Martin SEKERÁK, Daniel FÁBRY 320

Steganalýza farebných statických obrazov na báze DWT transformácie
Branislav GALLIK, Gabriel BUGÁR, Vladimír BÁNOCI, Dušan LEVICKÝ 326

Digitálna vodotlač vo videu na báze hybridného algoritmu DCT-SVD
Peter GOČ-MATIS, Martin BRODA, Dušan LEVICKÝ 330

Typy a aplikácie optických korelátorov
Tomáš HARASTHY, Ľuboš OVSENÍK, Ján TURÁN 337

Modul senzora teploty a vlhkosti s integrovaným rozhraním 1-Wire
Peter HUSNAJ, Miloš DRUTAROVSKÝ 343

MHP aplikácie v interaktívnej televízii
Rastislav KOKOŠKA, Stanislav MARCHEVSKÝ 349

Vyhladzovanie textúr 3R modelu ľudskej hlavy
Ondrej KOVÁČ, Ján MIHALÍK 355

Kombinovaná digitálna vodotlač v statických obrazoch
Zita KLENOVIČOVÁ, Ján CHALUPKA, Dušan LEVICKÝ 361

Návrh I-Q demodulátora pre UWB senzorový systém na báze M-postupností
Martin LIPTAJ, Matej ŽIGA, Pavol GALAJDA 365

A/D prevodník pre UWB senzorové systémy
Martin LIPTAJ, Tomáš HRAŠOK, Pavol GALAJDA 370

Design Texture Basis of Karhunen-Loeve Transform of Human Face
Ján MIHALÍK 376

Dostupnosť hybridnej FSO/RF linky
Pavol MIŠENČÍK, Ján TURÁN, Ľuboš OVSENÍK 381

Hybridný FSO/MIMO RF prenosový systém
Pavol MIŠENČÍK, Ján TURÁN, Ľuboš OVSENÍK 387

Vlastnosti hybridnej FSO/RF linky so 60 GHz RF záložnou linkou
Pavol MIŠENČÍK, Ľuboš OVSENÍK, Ján TURÁN 393

Prepínanie RF a FSO prenosovej linky hybridného FSO/RF systému
Pavol MIŠENČÍK, Ľuboš OVSENÍK, Ján TURÁN 399

Slepá steganalýza v statických obrazoch
Daniel MAJERČÁK, Vladimír BÁNOCI, Gabriel BUGÁR, Dušan LEVICKÝ 405

Distribuovaný dialógový manažér pre multimodálny dialógový systém
Stanislav ONDÁŠ, Martin VIŠŇOVSKÝ, Matej MOROZ 410

Optický vláknový refraktometer pre meranie indexu lomu kvapalín
Ľuboš OVSENÍK, Ján TURÁN 415

Optický vláknový refraktometer riadený cez internet
Ľuboš OVSENÍK, Ján TURÁN 421

Analógový pásmový ekvalizér s digitálnym riadením
Marek PETRINEC, Miloš DRUTAROVSKÝ 427

Sledovanie osôb za stenou pomocou impulzového UWB radaru
Peter PETRUŠKA, Jana ROVŇÁKOVÁ 433

ECC Cryptographic Library for ARM Processors
Martin PETRVALSKÝ, Miloš DRUTAROVSKÝ 439

Detekcia chybnej vzorky a jej automatická korekcia pri meraní polohy sledovaného mobilného objektu
Matúš PLEVA, Anton ČIŽMÁR 445

Využitie IP telefónie pre služby rečového komunikačného servera a jeho dodatočné zabezpečenie
Matúš PLEVA, Anton ČIŽMÁR 450

Modely dôvery pre oportunistické siete
Ján PAPAJ, Ľubomír DOBOŠ, Peter MALIC 455

Teória hier a jej využitie pre bezpečnosť v kooperatívnych sieťach
Ján PAPAJ, Ľubomír DOBOŠ, Tomáš LOZIAK 459

Three-dimensional OFDM in Frequency-selective Fading Channel
Ján PASTIRČÁK, Denis DUPÁK, Dušan KOCUR, Juraj GAZDA 465

Farebné filtre v predspracovaní obrazu pre optický korelátor
František RAKOCI, Ján TURÁN, Ľuboš OVSENÍK 471

Inventarizačný systém dopravných značiek s využitím optického korelátora
František RAKOCI, Ľuboš OVSENÍK, Ján TURÁN 477

Testovanie precíznych ČAP založené na porovnávaní s testovacím napätím
Martin SEKERÁK, Linus MICHAELI, Ján ŠALIGA, Jozef LIPTÁK, Marek GODLA 482

Bandpass Sigma-Delta ADC implemented on PSoC and MSI circuits
Martin SEKERÁK, Linus MICHAELI, Ján ŠALIGA, Marek GODLA, Jozef LIPTÁK 487

Analýza morfémových modelov slovenského jazyka v systémoch rozpoznávania plynulej reči
Ján STAŠ, Daniel ZLACKÝ, Jozef JUHÁR 494

Krátky vstup do problematiky analýzy chýb v systémoch rozpoznávania reči v slovenčine
Ján STAŠ, Pavol KOŠČ, Jozef JUHÁR 500

Morfologicky-založené modely slovenského jazyka pre systémy rozpoznávania plynulej reči
Ján STAŠ, Tibor ROVENSKÝ, Jozef JUHÁR 506

Evaluácia difónovej syntézy reči z textu slovenského jazyka
Martin SULÍR, Jozef JUHÁR 512

Programový modul funkcie predpovede počasia pre interaktívny rečový komunikátor
Marcel STRUCKEL, Matúš PLEVA 518

Systému na zber a analýzu analógových fyziologických signálov snímaných z mozgu a miechy pokusných zvierat
Martin SZARKA, Ján ŠALIGA, Ján GÁLIK 524

Comparison of Maximum Likelihood and Least Squares Fitting for ADC Testing in Simulated Environment
Ján ŠALIGA, Linus MICHAELI, Ján BUŠA, Jozef LIPTÁK, Marek GODLA 529

Senzor hmly na LOS KEMT FEI TUKE
Matúš TATARKO, Ján TURÁN, Ľuboš OVSENÍK 536

Dostupnosť a spoľahlivosť FSO linky vzhľadom na viditeľnosť
Matúš TATARKO, Ľuboš OVSENÍK, Ján TURÁN 540

Detekcia vodorovného značenia pomocou optického korelátora
Ján TURÁN, Ľuboš OVSENÍK, Tomáš HARASTHY 546

Fepstrálna analýza v systéme automatického rozpoznávania reči
Peter VISZLAY, Miroslav MUCHANIČ, Jozef JUHÁR 550

Banky filtrov pre implementáciu diskrétnej multiwaveletovej transformácie
Jozef ZAVACKÝ, Ján MIHALÍK 554

Celočíselná diskrétna waveletová transformácia
Jozef ZAVACKÝ 560

Zmena vzorkovacej frekvencie s neceločíselným faktorom
Jozef ZAVACKÝ 565

Design of Solution for Video Streaming of Interactive Whiteboard Surface
Dávid CYMBALÁK, Miroslav MICHALKO 571

Category for Component-based Program System
Marián JENČIK, Daniel MIHÁLYI 575

Visualization of PCB Fabrication in Education
Ondrej KAINZ, Slavomír KARDOŠ 580

Recording of Information on VoIP Calls to Calendar System at DCI
Katarína KLEINOVÁ, Peter FECIĽÁK, Michal BENČKO 585

GPGPU Technology Use for Security Solving in Network Environment
Jozef KOKORUĎA, Branislav SOBOTA 590

How to Express an Action Semantics of Programming Languages in Categories
Matúš KOLESÁR, Viliam SLODIČÁK 594

3D Engines for Visualization of Formally Developed Control Systems
Štefan KORENČKO, Peter NATAFALUŠI 601

Drawing System Using Virtual-Reality Technologies
Tomáš KOŠÍK, František HROZEK, Branislav SOBOTA 605

Simple Protection of VoIP Networks Against SPIT Attacks
Lukáš LAJOŠ, Jozef JANITOR 610

The Role of Digraphs in Modelling Processes
Martina ĽAĽOVÁ, Valerie NOVITZKÁ 614

The Rôle of Designs in Linear Logic
Pavol MACKO, Valerie NOVITZKÁ, Viliam SLODIČÁK 620

How to Express Denotational Semantics
Tomáš MARTINKO, Viliam SLODIČÁK 624

Time-Spatial Manipulation about Intrusion Detection System Knowledge
Daniel MIHÁLYI 630

System for Support of Adaptive Selection of Holiday Destination
Zuzana PANCÁKOVÁ, Viliam SLODIČÁK 635

The SLA Meter Tool
Adrián PEKÁR, Juraj GIERTL, Martin RÉVÉS, Peter FECIĽAK, Miroslav ANTL 641

The Rôle of Abstract Data Types and Abstract Behavior Types
Marián POLÁK, Viliam SLODIČÁK 647

3D Training Manual with Augmented Reality Technologies
Monika RAŠKOVSKÁ, Branislav SOBOTA, Martin VARGA 651

Geographical Information Systems and 3D Visualization Technologies
Branislav SOBOTA, František HROZEK, Lukáš FEDORKO 655

An Instruction Decoder and Disassembler Generator for EmuStudio Platform
Matúš SULÍR, Slavomír ŠIMOŇÁK 660

Distributed Cloud-Centric System Call Tracing Reputation Based (DCSR) Intrusion Prevention System (IPS)
Martin TOMÁŠEK, Marek ČAJKOVSKÝ, Ivan KLIMEK 664

Electric Field Influence on Steady Element of Charging Current for High Voltage Insulation Material
Roman CIMBALA, Martin GERMAN-SOBEK, Jozef KIRÁLY 668

Modely prevádzok vn elektrizačných sústav a ich porovnanie z hľadiska zavlečenia prepätí do nn rozvodov
Bystrík DOLNÍK, Dominik SMOLKO 674

IRC Analysis of Degraded XLPE Cable
Martin GERMAN-SOBEK, Jozef KIRÁLY, Roman CIMBALA 679

Konštrukcia modelu veterných elektrární
Slavomír HERBERT, Ján TKÁČ 685

Využitie SVC na minimalizáciu činných strát v elektrizačnej sústave
Roman JAKUBČÁK, Ľubomír BEŇA 691

Návrh a konštrukcia ovládacej jednotky pre polohovacie zariadenie
František JUSKO, Ján TKÁČ 695

Transformer Oil Based Ferrofluid From View of i(t) Characteristics
Jozef KIRÁLY, Martin GERMAN-SOBEK, Roman CIMBALA 701

ISBN 978-80-553-0890-6 © 2012 FEI TUKE

Skúšky digitálnych ochrán prístrojom OMICRON CMC
Miroslav KMEC, Marek HVIZDOŠ..... 705

Viskozita prírodných esterov
Iraida KOLCUNOVÁ, Lýdia DEDINSKÁ, Viera VANČOVÁ..... 711

Štatistická analýza čiastkových výbjov
Iraida KOLCUNOVÁ, Marián HRINKO, Juraj KURIMSKÝ..... 717

Meranie meteorologických parametrov pri využívaní slnečnej energie
Peter KOLPASKÝ, Ján TKÁČ..... 722

Komunikácia digitálnych ochrán pomocou protokolu IEC 61850
Andrej MASAROVIČ, Marek HVIZDOŠ..... 728

Využitie nízkopotenciálneho tepla pomocou tepelných čerpadiel
Peter MASKAĽ, Marek HVIZDOŠ, Ján TKÁČ..... 734

Modelovanie elektromagnetických polí v okolí stožiarov vvn a zvn
Dušan MEDVEĎ, Jozef ŠURIN..... 740

Modelovanie prechodných javov využitím nástroja Matlab/SimPowerSystem
Dušan MEDVEĎ, Lukáš NEMERGUT..... 745

Príspevok k bezdrôtovému prenosu elektriny
Matej ORUŽINSKÝ, Juraj KURIMSKÝ..... 749

Optimal Design of a Fuzzy DC Drive Controller
Pavol FEDOR, Daniela PERDUKOVÁ..... 754

Neural Estimators for Shaft Sensorless FOC of Induction Motor
Peter GIROVSKÝ, Jaroslav TIMKO, Jaroslava ŽILKOVÁ..... 760

Simulácia meničových systémov v prostredí OrCAD/PSpice a MATLAB/Simulink
Milan LACKO, Peter GIROVSKÝ..... 767

Využitie ANFIS štruktúry pre odhad rýchlosti AM
Peter NGUYEN, Jaroslava ŽILKOVÁ, Marek VACEK..... 773

Riešenie inverznej kinematickej úlohy sériového robotického ramena pomocou neurónových sietí
Marek VACEK, Jaroslava ŽILKOVÁ, Peter NGUYEN..... 780

Concept of the Virtual Laboratory for the Assembling Technologies in Electronics
Dominik DEMETER, Juraj BANSKÝ..... 786

Design and Comparison of Different Methods of Measurement of Low Pass Filter for Ultra Wide-Band Application
Alena PIETRIKOVÁ, Kornel RUMAN, Igor VEHEC, Pavol GALAJDA..... 789

Simulation of Stencil Printing Process
Alena PIETRIKOVÁ, Michal KRAVČÍK..... 795

Influence of Ultrasonic Wedge Bonding Parameters on Resistivity
Igor VEHEC, Alena PIETRIKOVÁ, Kornel RUMAN..... 800

Intricate World of Electrical, Magnetic and Structural Properties of Rare Earths Metals Thin Films at Low Temperatures – the Case of Holmium and Thulium Thin Films
Ján DUDÁŠ, Jozef BAGI, Milan GUZAN, Slavomir GABÁNI 805

Diagnostika napájacej sústavy automobilu na báze internetu
Ján MOLNÁR, Tibor VINCE 811

Vývoj softvéru pre vzdialené meranie variabilných topológií
Tibor VINCE 818

The Crossing Number of $G \square P_n$ for the Specific Graph G of Order Six
Marián KLEŠČ, Daniela KRAVECOVÁ, Matúš VALO 822

Weather Prediction Using Neural Networks

Ján ADAMČÁK, Rudolf JAKŠA

Department of Cybernetics and Artificial Intelligence, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Slovak Republic

j.adamcak@gmail.com, jaksa@neuron.tuke.sk

***Abstract* — This paper deals with a weather prediction in the region of Košice – Technical University. Data, which are used in experiments, were processed by Ing. Juraj Koščák via the WeatherUnderground, Inc. for Košice – Airport place. In this project we use a Weather Station Davis Vantage Pro 2, whose is installed in the premises of the Technical University of Košice. This paper describes how to build a system for collection, prediction and presentation of meteorological data in the area of the Technical University in Košice. This work also includes the evaluation of experimental learning neural network for prediction of meteorological data. The experiment is to build a functional system for the collection, prediction and presentation of meteorological data.**

***Keywords* — Weather prediction, Backpropagation, Neural Networks, Weather Station**

I. INTRODUCTION

This work deals with a weather prediction in the region of Košice – Technical University. This paper is a summary of master thesis, Technical University of Košice 2012 [1].

Weather is the state of the atmosphere, to the degree that it is hot or cold, wet or dry, calm or stormy, clear or cloudy [3]. Most desired services related to the weather is weather forecast. Main part of weather forecast is prediction of temperature.

Weather forecast is the prediction generated by human from output of mathematical-physical model of the atmosphere. This model is computed on supercomputers. This method of computing future state of atmosphere is called numerical. Other methods of generating forecast are statistical methods, synoptic method and climatology method.

In this paper is used one of statistical methods. This method is called neural networks method. In this method are used neural networks with different topology. Main topology to work with time series is recurrent neural networks. These networks have an internal memory. From these networks are derived feed-forward neural networks, where is internal memory simulated a time window with more inputs neurons. This time window is replacement for the internal memory.

During research publicated in this paper was used feed-forward neural network with one hidden layer. This network has implemented backpropagation algorithm. Purpose of project was learning of meteorological rules from historical data (10 years old) and use this rules for local weather prediction. This network is generating prediction for next 24 hours every 15 minutes after measurement by weather station.

II. ARCHITECTURE OF SYSTEM FOR MEASUREMENT, PREDICTION AND VISUALIZATION OF METEOROLOGICAL DATA

A. Scheme of system

System for measurement, prediction and visualization of meteorological data is divided into the following sections: measurement section, prediction section and visualization section. These sections are synchronized via crontab. Crontab is a Linux utility for scheduling tasks. Standard task for this system has this advisory role: make a measurement, write measured data to archive, generate time window for predictor, run predictor, edit output of predictor, visualize measured data and visualize predicted data.

Scheme of system for measurement, prediction and visualization of meteorological data is on **Fig. 1**.

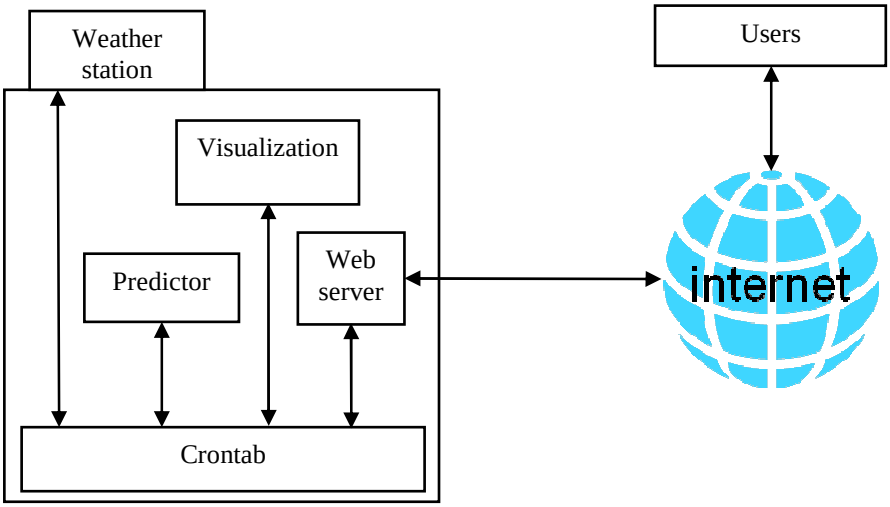


Fig. 1 Scheme of realized system for measurement, prediction and visualization of meteorological data.

B. Measurement section

Measurement section of system is divided to Weather station Davis Vantage Pro 2 and free Linux driver for this Weather station. Weather station is installed in Technical University of Košice area on the roof of Center for Intelligent Technologies. For communication between Weather station and personal computer was edited driver from Swinburne University of Technology, Melbourne. This driver was published with GPL license in programming language C.

C. Prediction section

Prediction section of system is neural network publicated in [2] and edited in [1]. This neural network use meteorological rules learned from meteorological data during experiments. These rules are written as weights for this neural network. Input of this neural network is generated as a time window of measurements from last 3 days. These measurements are divided into measurement of outside temperature, dew point, humidity, barometer and wind speed and direction. Output of this neural network is prediction of next 48 measurements (24 hours). This prediction section is running after measurement section.

D. Visualization section

Visualization section of system is group of GNUPLOT scripts. These scripts generate graphical visualization of meteorological variables. These variables are parsed to 3 day time window, similarly as in prediction system. Visualization section generates graphs of outside temperature, humidity, barometer, wind speed and direction, precipitation and solar radiation. Visualization section generates prediction graphs, too. The generated graphs are sent to webpage for publication on webpage of project. This webpage has an address <http://neuron.tuke.sk/~adamcak/0/eng.html>.

III. EXPERIMENTS

During working on this research we realized these experiments: experiments of learning main neural network on data from Košice-Airport, experiments of prediction for Technical university of Košice area and experiments of relearning neural networks with data from weather station installed in Technical University of Košice.

A. Experiments of learning of main neural network

During experiments of learning main neural network we used data from Weather Underground Inc. In these data we measured last 10 years for Košice-Airport area. Learning progress of one of experiments of learning is visible on **Fig. 2**. In this experiment γ parameter was set to 0.0003. Graphical analysis of medium square error (MSE) during learning of 10 years of data on training set and testing set did show that best weights vector is in 2737 iteration of learning cycle. This weights vector is used for prediction in experiments with predictions.

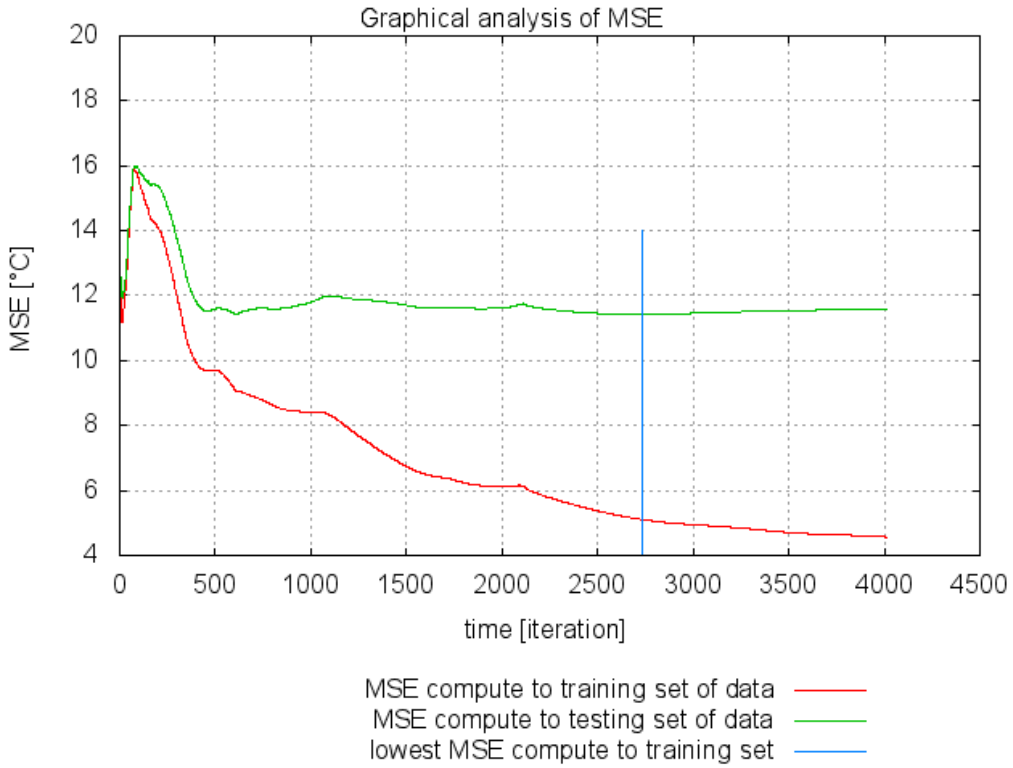


Fig. 2 Graphical analysis of MSE during experiments of learning 10 years weather history in Košice-Ariport. Parameter γ=0.0003.

B. Prediction experiments

During prediction experiments we used the best weights vectors from experiments of learning main neural network. One result of experiments of prediction is displayed on **Fig. 3**. This prediction was generated at 25th April 2012 at 0:00.

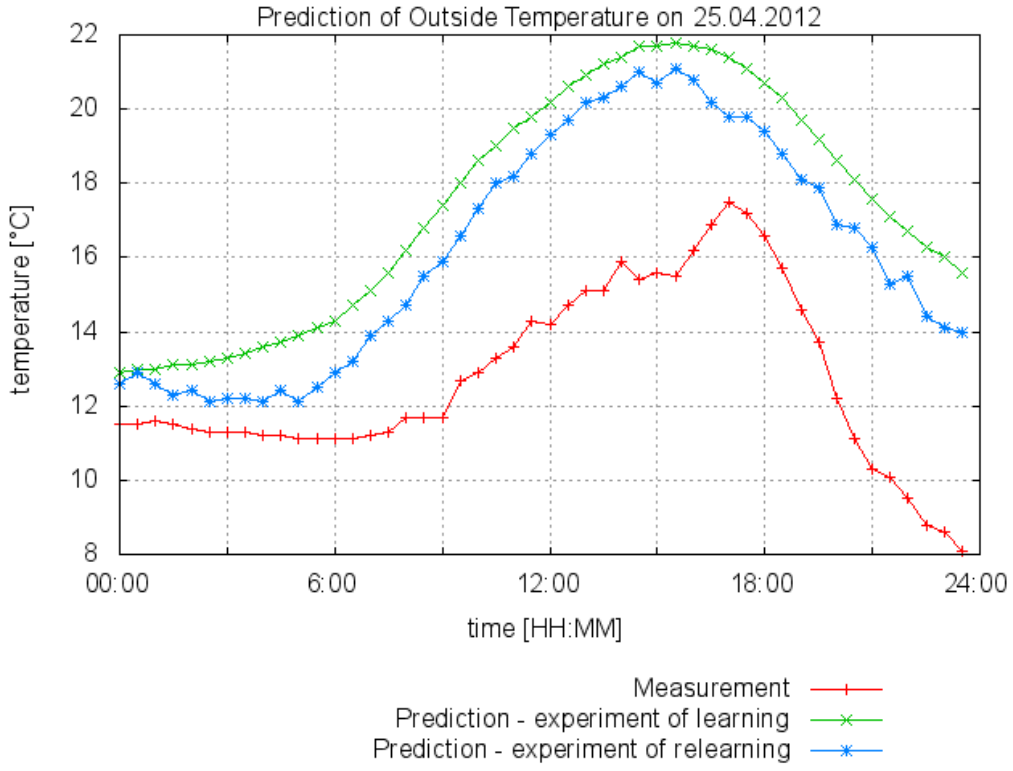


Fig. 3 Graphical analysis of outside temperature measured and predicted for Technical University of Košice on 25th April 2012.

C. Experiments with relearning

Experiments of relearning have not produced the excepted improvement in results of prediction. These experiments used data from weather station installed in Technical University of Košice area and neural networks initialized by weight vectors from experiments of learning main neural networks. These experiments of relearning were tested in two points of relearning. First point was after relearning in 2000th iteration of relearning. Second point was in change a speed of error of relearning. Results of relearning experiments are published in Table 1. One reason of not so good results in this series of experiments is short history measured by our weather station. This history is long only 10 months. The best network in **Table 1** is network learned in exp.3 and relearned in rel.3 and rel.3p. The worst network in **Table 1** is network learned in exp.2 and relearned in rel.2 and rel.2p. Experiment rel.1 and rel.1p is not relearned network from exp.1. Experiment rel.1 and rel.1p learn 10 month data from our weather station on the main neural network.

Table 1
Results of experiments. Rows are summed and average error of prediction at 25th April 2012. Culloms are experiments of learning and relearning. Culloms exp.1, exp.2 and exp.3 are results of experiment of learning 10 years history on the main neural network. Culloms rel.1, rel.2 and rel.3 are results of experiment of relearning from 2000th iteration. Culloms rel.1p, rel.2p and rel.3p are results of experiment of relearning from point of change the speed of error.

	Experiments of learning			Experiments of relearning			Experiments of relearning		
	exp.1	exp.2	exp.3	rel.1	rel.2	rel.3	rel.1p	rel.2p	rel.3p
sum	41,7	53	225,3	292,9	285,8	126,5	167,5	177,7	104,5
avg	0,87	1,10	4,69	6,10	5,95	2,64	3,49	3,70	2,18

IV. CONCLUSION

System build in this research is working online in real-time and generating predictions from 1st May 2012. This system works well, when weather is stable. Because system has only local data, prediction is not good, when sudden change in the weather.

Another weak point in our predictor was missing information about clouds on input of predictor. Clouds cause a little change of temperature. In future, to raise quality of prediction we need better quality and quantity meteorological data from weather station installed in this system.

ACKNOWLEDGMENT

This work is the result of the project implementation: Development of the Center of Information and Communication Technologies for Knowledge Systems (ITMS project code: 26220120030) supported by the Research & Development Operational Program funded by the ERDF.

REFERENCES

[1] J. Adamčák, Weather Prediction Using Neural Network, Master's thesis, Technical University of Košice, Faculty of Electrical Engineering and Informatics, 2012, 57 p.
Online: http://neuron.tuke.sk/~adamcak/theses/Adamcak-Jaksa-MS-thesis-2012.pdf

[2] J. Adamčák, Weather Prediction Using Neural Network, Bachelor's thesis, Technical University of Košice, Faculty of Electrical Engineering and Informatics, 2010, 54 p.
Online: http://neuron.tuke.sk/~adamcak/theses/Adamcak-Jaksa-BC-thesis-2010.pdf

[3] Merriam-Webster Dictionary: Weather. In: [online]. [cit. 2012-05-10]. See online: http://www.merriam-webster.com/dictionary/weather

Riadenie robota rozpoznávaním ľudských pohybov pomocou zariadenia Kinect

Peter ADAMONDY, Peter PAPCUN, Ján JADLOVSKÝ

Katedra kybernetiky a umelej inteligencie, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

michal.sakmar@tuke.sk, gabriel.bugar@tuke.sk, jan.jadlovsky@tuke.sk

***Abstrakt* — Predkladaná práca pojednáva o návrhu a vytvorení modulárnej aplikácie riadenia robota rozpoznávaním ľudských pohybov pomocou senzora kinect. Riadiť budeme robotickú ruku RV-2SDB od spoločnosti Mitsubishi, ktorá je v súčasnosti integrovaná v pružnej výrobnej linke. Navrhneme nový prístup k vytváraniu pracovných sekvencií, pričom na ovládanie použijeme ruku operátora, ktorej pohyb bude zaznamenaný a použitý ako predloha trajektórie robotického ramena.**

***Kľúčové slová* — robot, riadenie, kinect, rozpoznávanie ľudských pohybov**

I. ÚVOD

Je prirodzené, pri riešení akejkoľvek úlohy využiť najmodernejšie postupy a technológie. Ak sa pozrieme na historicky vývoj, ktorým si svet technológii prešiel, nemôžu nám uniknúť zmeny v prístupoch človeka k nástrojom vlastnej prace. Je v ľudskej prirodzenosti vytvárať nástroje, ktoré uľahčujú a zrýchľujú jeho činnosť. Najmä v moderných ergatických systémoch si však často kladieme otázky súvisiace s interakciou medzi človekom a strojom. Preto vytvoríme systém riadenia robota, tak aby bol schopný kopírovať pohyby užívateľa. Aplikácia umožni tieto pohyby uložiť a použiť ich pre programovanie rutinných sekvencií, ktoré robot ma vykonávať. Takýto prístup dovolí extrémne rýchle preprogramovanie robota na novu úlohu pri minimálnom úsilí. Tiež je možne takéto riadenie použiť na operácie, ktoré sa veľmi zložito programujú, alebo priamo vyžadujú účasť obsluhy na vykonávanej úlohe, ktorá však nesmie, či nemôže fyzicky vkročiť do miesta samotného prevádzania operácie.

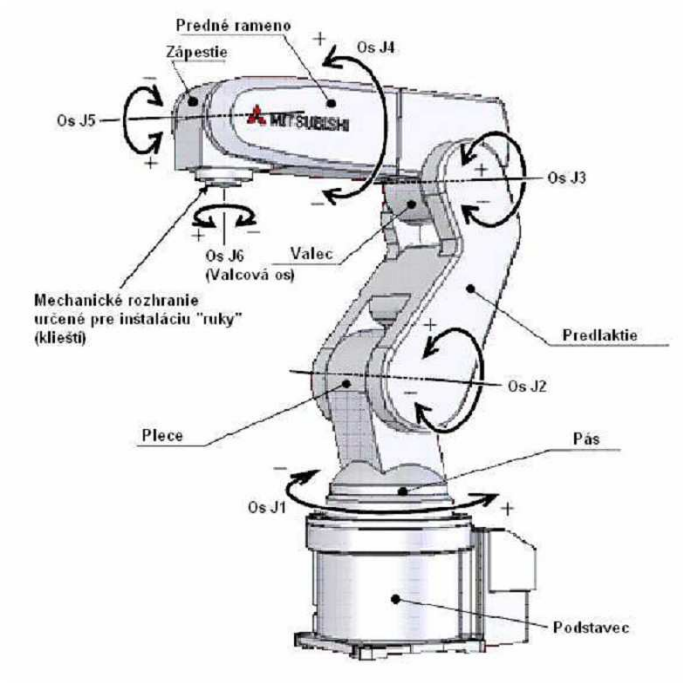
II. ROBOT MITSUBISHI RV-2SDB

Mitsubishi RV-2SDB je kompaktný, modulárny robot využívaný v priemyselných výrobných linkách na rôzne operácie (presúvanie materiálu, rezanie, lakovanie atď.). Ma šesť stupňov voľnosti, s maximálnou váhou bremena 3kg. Fotografia robota je na Obr. 1. Nachádza sa v laboratóriu V147 Katedry kybernetiky a umelej inteligencie, kde je súčasťou pružnej výrobnej linky. Viac v [3].



Obr. 1 Robotické rameno Mitsubishi RV-2SDB

Riadenie zabezpečuje riadiaca jednotka Mitsubishi CR1D, s ktorou PC komunikuje pomocou Ethernet pripojenia. Samozrejme, riadiaca jednotka dokáže pracovať aj s inými sieťovými rozhraniami (ProfiBus, RS-232)



Obr. 2 Osi robota Mitsubishi melfa RV-2SDB

III. KINECT

Kinect je špeciálne vstupne zariadenie pôvodne určené pre hernú konzolu Microsoft Xbox 360. Jeho úlohou je snímať pohyby tela hráča a prenášať ich do herného prostredia. Na trh sa Kinect dostal v novembri 2010. Vzhľadom k veľkému záujmu využitia tejto technológie aj k iným účelom než k hraniu hier, bol v júni 2011 vypustený Kinect SDK (Software Developer Kit). Jednalo sa o balíček programátorských nástrojov slúžiacich na integráciu kinectu do užívateľských aplikácii.

Aj keď bol vyvinutý spoločnosťou Microsoft, jeho najdôležitejšia časť pochádza od izraelskej firmy PrimeSense, ktorá vyvinula technológiu ovládania elektronických zariadení pomocou infračerveného projektora a snímača. Túto technológiu nazvala LightCoding.



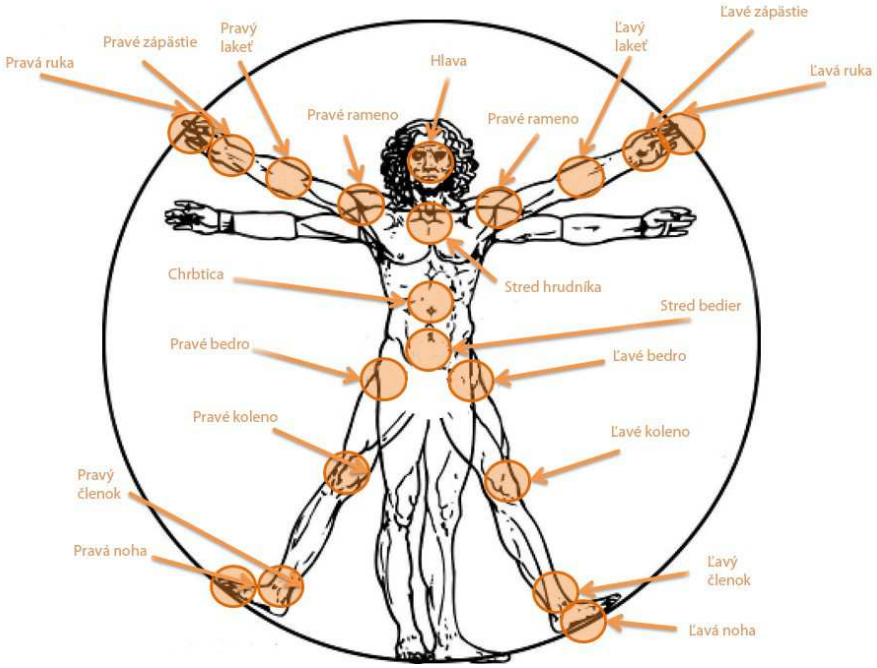
Obr. 3 Rozloženie jednotlivých prvkov zariadenia Kinect

Projektor vysiela do priestoru infračervené lúče. Senzor potom zachytáva ich odraz od povrchu. Intenzita odrazeného infračerveného žiarenia v kombinácii so vzájomnou vzdialenosťou týchto bodov vytvára hĺbkovú mapu.

IV. ROZPOZNÁVANIE KOSTRY

Rozpoznávanie kostry využitím hĺbkovej mapy prináša niekoľko výhod oproti tradičným technikám rozpoznávania z farebných obrázkov. Môžu pracovať pri akýchkoľvek svetelných podmienkach, bez ohľadu na farbu či textúru. Zjednodušujú pracovanie obrazu a tým aj odstraňovanie pozadia. Najvýznamnejšou výhodou je jednoduchá extrakcia hĺbkovej siluety užívateľa, čo umožňuje ľahké získavanie trénovacích dát. Algoritmus detekcie kostry a pohybu bol navrhnutý týmom na Cambridgeskej univerzite v spolupráci so spoločnosťou Microsoft [1]. Syntetické dáta získané z algoritmu používajúcom motion capture dáta, vytvára hĺbkovú mapu renderovaním 3D modelov postav. Takto získané dáta slúžia na trénovanie rozhodovacích stromov zlúčených do rozhodovacieho lesa. Pred tým než sa však dajú použiť, musia sa vymedziť časti tela užívateľa. Ide o definovanie zón na povrchu siluety postavy. Niektoré z týchto zón predstavujú oblasti v ktorých sa nachádzajú kĺby, iné vypĺňajú priestor medzi nimi. Silueta je v algoritme rozdelená na 31 oblastí, z toho 20 oblastí predstavuje možnú polohu kĺbu. Iba kvôli presnosti je nutne podotknúť, že pojem „kĺb“ predstavuje všetky body ohybu kostry a teda nie len fyzické kĺby užívateľa.

Výsledkom algoritmu sú súradnice dvadsiatich kĺbov v 3D priestore. Tieto kĺby, resp. „body ohybu“ sú znázornené na Obr. 4



Obr. 4 Polohy detekovaných kĺbov

V. NÁVRH ARCHITEKTÚRY APLIKÁCIE

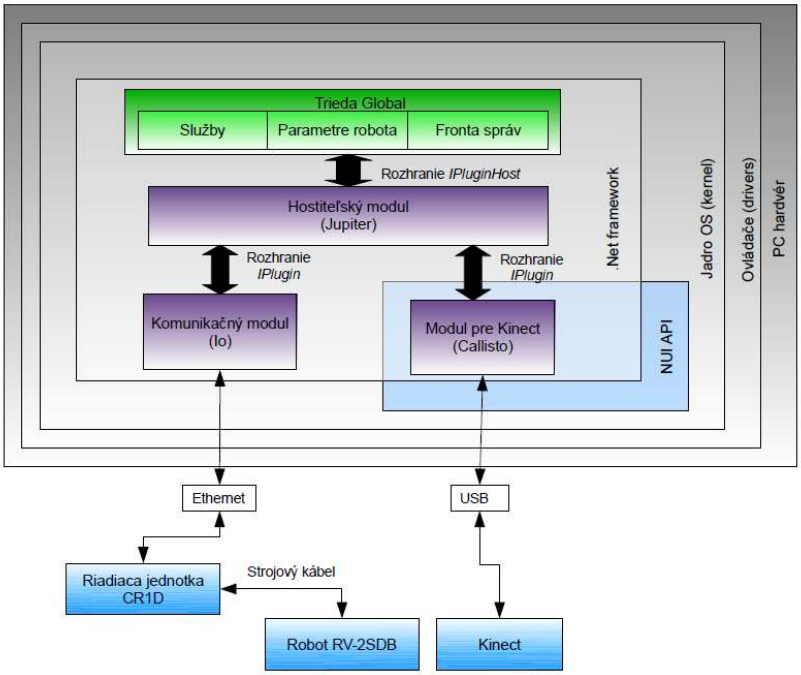
Ak chceme vytvoriť moderný riadiaci systém, musíme vytýčiť základné piliere dizajnu, resp. požiadavky na funkcionalitu. Keďže reálne využitie radenia robotov pomocou rozpoznávania pohybov ešte ani nevieme plne predvídať (vytváranie automatizovaných sekvencií, operácie pacientov na diaľku, presun a manipulácia materiálu v neprístupných miestach atď.), musí byť systém dostatočne univerzálny, aby sa vedel prispôsobiť čo najširšiemu spektru úloh. Je tu teda požiadavka modularity riešenia.

V praxi dochádza často k situácii kedy je nutné meniť systémy riadenia počas samotnej prevádzky. Možnosť pridávať, či upraviť funkcionalitu systému, bez nutnosti zastavenia prevádzky, môže nie len uľahčiť implementáciu nových častí, ale aj ušetriť náklady spojené s odstávkou celého systému. Definujme teda potrebu meniť vlastnosti aplikácie za behu, ako ďalšiu požiadavku.

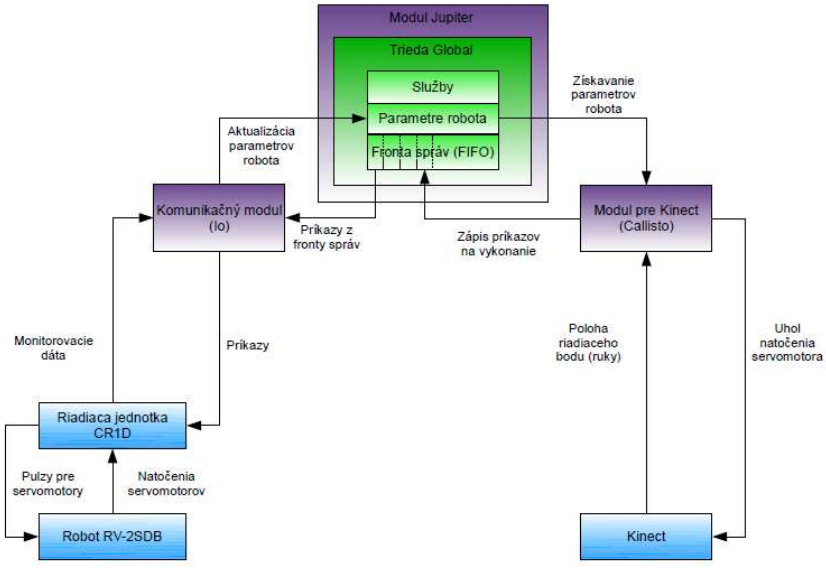
Z hore spomenutých podmienok dizajnu vyplýva potreba použiť princípy modulárneho programovania. Modulárne programovanie nie je žiadnou novinkou. Jeho implementáciu nájdeme v mnohých moderných informačných systémoch, aplikáciách, či aplikačných serveroch. Modulárny systém je taký, ktorého funkcie sú tak povediac zabalené do modulov (V angličtine sa dajú nájsť aj pod pojmom „plug-ins“, preto sa v aplikácii odkazujem na moduly ako na pluginy). Každý z týchto modulov má v rámci systému svoju úlohu a skladá sa z dvoch logických častí. Prvou časťou je samotný vnútorný systém modulu, ktorý sa môže od ostatných

líšiť *de facto* ľubovoľne. Druhá časť je vonkajší interfejs. Teda rozhranie ktorým modul komunikuje s vonkajškom. Interfejs funguje ako „lepidlo“ medzi modulmi.

Pre realizáciu návrhu architektúry celého systému sme museli vyriešiť, akým spôsobom bude nami vytvorená aplikácia komunikovať s hardvérovými prvkami. Aplikácia využíva .Net framework, ako skupinu štandardných knižníc nad operačným systémom. Modul Callisto okrem toho k získaniu dát implementuje aj niektoré prvky NUI API (Natural User Interfaces). Vďaka nemu môžeme pristupovať na senzor kinect. Globálna schéma systému vrátane komunikačných rozhraní je na Obr. 5.



Obr. 5 Globálna architektúra systému



Obr. 6 Komunikačný dátový model systému

Model na Obr. 6 reprezentuje kolobeh dát v rámci systému. Ako vidíme, sú tu dva uzavreté obvody riadenia. Prvá spätnoväzobná slučka (vnútorný obvod) sa skladá z regulátorov v riadiacej jednotke CR1D a servomotorov v robotovi. Druhý (vonkajší) riadiaci obvod sa skladá z modulu Callisto a riadiacej jednotky, ako riadeného systému. Vonkajší riadiaci obvod generuje trajektóriu (ako sled príkazov posielaných modulom Io) pre riadiacu jednotku robota. Treba poznamenať, že vnútorná slučka je z vonku neprístupná, t.j. nemôžeme ju nijako meniť (bola vytvorená spoločnosťou Mitsubishi).

Treba si uvedomiť, že sa tu stretávame s dvoma typmi modulov. Jeden predstavuje hostiteľa a druhý je hosťom. Zatiaľ čo hostiteľ je vždy len jeden, hostí môže byť viac. Hostia využívajú služby hostiteľa a on ich informuje o rôznych udalostiach. Pre jasnejšie vysvetlenie si predstavme nasledujúci príklad. Užívateľ sa rozhodne zmeniť jazyk aplikácie. Každý modul v sebe cez interface implementuje metódu na zmenu jazyka s názvom „zmenJazyk()“. Hostiteľský modul upozorní všetky ostatné moduly tým, že túto metódu zavolá. Výhodou použitia interface [2] je práve nutnosť implementácie všetkých metód, ktoré sú v ňom definované. Hostiteľ sa teda nemusí obávať, že niektorý modul túto metódu nemá.

VI. MEDZIMODULOVÁ KOMUNIKÁCIA

Moduly sú samostatné objekty s vlastnou vnútornou logikou. Navonok sa však musia chovať uniformne. Keďže moduly majú svoje úlohy, musí byť zabezpečená vzájomná komunikácia. Všetky informácie budú prechádzajú hostiteľským modulom, avšak ho nijako nesmú zaťažovať. Môžeme použiť metaforu s hostiteľom. Najlepším riešením by bola možnosť vymeniť si informácie tak, že jeden hosť zapíše do hostiteľskej „knihy návštev“ informáciu a druhý hosť si ju vyčíta keď potrebuje. V podstate hostiteľ v tejto komunikácii hrá rolu „strážcu knihy“. V rámci systému bude dostupná fronta správ pre všetky moduly, ktorýkoľvek do nej môže zapisovať a vyberať informácie podľa potreby. Fronta je generická, t.j. správy sa môžu líšiť. V našom prípade budeme posielať príkazy z modulu riadenia (Callisto) do komunikačného modulu Io. Io bude zas zodpovedný za aktualizáciu informácií o robotovi.

VII. ZÁVER

V tejto práci sa ukázalo, akým spôsobom sa dajú využiť pohyby človeka na ovládanie robotického ramena. V budúcnosti by sa tento, alebo podobný systém, dal využiť pri operáciách pacientov. Robot môže nahradiť ruku človeka v miestach, do ktorých sa operátor, či užívateľ nemôže dostať. Ďalšie využitie vidím v rýchlom preprogramovaní robotov vo výrobných linkách. Čitateľa iste napadnú vlastné možnosti aplikácie.

Zhrňme si teda, čo sa nám v tejto práci podarilo. Analyzovali sme možnosti rozpoznávania pohybov človeka pomocou kinectu. Detekciou a následnou transformáciou polohy ruky užívateľa sme vytvoril mechanizmus riadenia v rámci modulárnej aplikácie. Návrh tejto aplikácie umožňuje počas chodu vypínať a zapínať moduly podľa potreby. Ak by sme napríklad chceli pridať riadenie nového robota, jednoducho by sme vytvorili nový modul, ktorý by komunikoval s riadiacou jednotkou, pričom by ostatné moduly zostali rovnaké. Vieme si predstaviť rozšírenie tohto systému napríklad modulom, ktorý by na ovládanie používal joystick, myšku, trackball, alebo v podstate akékoľvek iné vstupné zariadenie. Návrh takéhoto typu riadiaceho systému nesporne prináša niekoľko výhod, keďže urýchli a zjednoduší vývoj nových riadiacich a monitorovacích aplikácií, čím ušetrí čas a finančné prostriedky.

POĎAKOVANIE

Táto práca bola vytvorená realizáciou projektu Rozvoj Centra informačných a komunikačných technológií pre znalostné systémy (kód ITMS projektu: 26220120030) na základe podpory operačného programu Výskum a vývoj financovaného z Európskeho fondu regionálneho rozvoja.

LITERATÚRA

[1] SHOTTON, J. – Fitzgibbon, A. et al.: Real-Time Human Pose Recognition in Parts from Single Depth Images, Microsoft Research Cambridge & Xbox Incubation, 2010. Dostupné na internete: http://research.microsoft.com/pubs/145347/BodyPartRecognition.pdf

[2] SHARP – JOHN: Microsoft Visual C# 2008. Brno : Computer Press, 2008, s.232-235. ISBN 978-80-251-2027-9

[3] PAPCUN, P.: Riadenie robota integrovaného v pružnej výrobnej linke. Diplomová práca. Košice: Technická univerzita v Košiciach, Fakulta elektrotechniky a informatiky, 2011. 95 s.

[4] ADAMONDY, P.: Riadenie robota rozpoznávaním ľudských pohybov pomocou zariadenia Kinect, Diplomová práca. Košice: Technická univerzita v Košiciach, Fakulta elektrotechniky a informatiky, 2012. 76s.

Využitie internetového marketingu pre zvýšenie predajnosti internetového obchodu

Martin ANDERKO, František BABIČ

Katedra kybernetiky a umelej inteligencie, Fakulta elektrotechniky a informatiky,
Technická univerzita v Košiciach, Slovenská republika

martin.anderko@student.tuke.sk, frantisek.babic@tuke.sk

***Abstrakt* — Internetový marketing predstavuje dôležitú možnosť ako zvýšiť predajnosť internetového obchodu. Na tento účel existujú viaceré metódy a prístupy ako SEO optimalizácia, reklamné PPC kampane, sledovanie návštevnosti a cielená reklama, atď. Tento článok popisuje implementáciu vytvoreného mixu internetového marketingu spolu s výsledkami reprezentujúcimi zvýšenie návštevnosti a počtu objednávok daného internetového obchodu. Zvolený mix predstavuje finančne nenáročné riešenie, ktoré môže byť inšpiráciou pre iné elektronické obchody alebo prezentačné stránky firiem.**

***Kľúčové slová* — internetový obchod, SEO, Google AdWords**

I. ÚVOD

Internetový marketing, PPC reklama a SEO sú pojmy, ktoré sa v posledným rokoch dostávajú do popredia stále viac. Tieto prístupy sú základom každého úspešného webového projektu. Podnikatelia už od nepamäti vyhľadávajú rôzne trhové príležitosti na generovanie ziskov z ich podnikateľskej činnosti a internet je nato v súčasnosti najlepším nástrojom. Firmy v súčasnosti zisťujú, že nie je dôležité len „mať internetovú stránku“, mnohé však nato stále zabúdajú. Niekedy nestačí len dobrý nápad, aj keď sa môže zdať, že z ničoho nič je tu nový projekt, ktorého tvorcovia sa stali cez noc boháčmi, avšak vždy za tým stojí mnoho času a investovanej energie.

Návod na úspešný internetový marketing firmy neexistuje, keďže každý webový projekt je špecifický. Môžeme však uvažovať o základných prístupoch, ku ktorým patria – kvalitné služby/produkty, dôraz na dizajn, prístupnosť, SEO (optimalizácia pre vyhľadávače), reklama, marketing na sociálnych sieťach, budovanie dôvery a udržiavanie komunikácie s klientmi. Základom úspešnej propagácie nie je masová reklama, ale ponúknuť budúcim zákazníkom to, čo hľadajú.

A. Elektronické obchodovanie

Elektronické obchodovanie sa využíva pre možnosť sprostredkovania a zabezpečenia služieb pre podnikateľskú sféru a bežných spotrebiteľov, ako aj pre ponuku priemyselných spotrebných predmetov cez webové stránky na internete. Zahŕňa priamy marketing od výrobcu, maloobchodníka alebo iných sprostredkovateľov ku konečnému používateľovi. Prakticky to znamená, že firma si zriadi domovskú webovú stránku na internete, uvádza na nej informácie pre nadviazanie kontaktu a internet využíva aj na predaj svojho tovaru. [1]

Pri tvorbe internetové obchodu si môžeme vybrať z dvoch alternatív: použitie voľne dostupného riešenia alebo vytvoriť systém na mieru. Obidve možnosti majú svoje výhody a nevýhody [2]. Obchody postavené na voľne šíriteľných systémoch sú v súčasnosti veľmi obľúbené. Medzi hlavné výhody týchto systémov patrí rýchlosť nasadenia systému a modularita súčasných systémov. Medzi najrozšírenejšie systémy na Slovensku patria Magento, PrestaShop a OpenCart. Spomenúť môžeme tiež systémy Zen Cart, OsCommerce, Drupal e-commerce a VirtueMart postavený na CMS systéme Joomla.

B. SEO optimalizácia

Skratka SEO vychádza zo začiatočných písmen „*search engine optimization*“, čo je možné voľne preložiť ako optimalizácia pre vyhľadávače. Proces optimalizácie webových stránok pre vyhľadávače je dlhodobý proces, ktorý si vyžaduje dobrú prípravu – dopady optimalizácie sa z pravidla prejavujú postupne. Podľa prieskumov predstavuje SEO najefektívnejší nástroj na získanie relevantného zákazníka. SEO pokrýva dva hlavné prístupy: on-page a off-page optimalizáciu. [3]

On-page optimalizácia sa sústreďuje hlavne na obsah, navigačnú štruktúru, titulky samotných stránok, ich prístupnosť a ďalšie faktory ovplyvňujúce priazeň vyhľadávačov. Táto časť optimalizácie sa často nazýva aj optimalizácia zdrojového kódu.

Za off-page faktory sú pri optimalizácií považované všetky faktory alebo úpravy, ktoré nie sú aplikované na konkrétnej internetovej adrese. Do off-page faktorov môžeme zaradiť napr. nákup spätných odkazov, registráciu do katalógov, výmenu odkazov, tvorba a publikácia PR článkov. Pri internetových obchodoch je taktiež vhodná registrácia do vyhľadávačov tovarov, ako napr. Heureka.sk, Najnakup.sk a podobne. Celý proces získavania kvalitných spätných odkazov sa nazýva linkbuilding[1].

C. PPC reklama

PPC reklama z anglického „*pay-par-click*“ znamená zaplať za klik. Znamená to, že inzerent za reklamu platí iba v tej chvíli, keď potenciálny klient na reklamu klikne. Takisto si inzerent vie sledovať, koľko za reklamu platí. Medzi najpoužívanejšie PPC systémy[2] v súčasnosti na Slovensku patria Google Adwords, Etarget a Facebook PPC. Google AdWords predstavuje PPC systém od Google, ktorý sa vyznačuje pomerne vysokými konverziami. Cena za klik sa v tomto systéme určuje podľa veľkosti konkurencie – čím viac firiem si platí reklamu na určité kľúčové slovo, tým vyššia je jeho cena.

II. INTERNETOVÝ OBCHOD

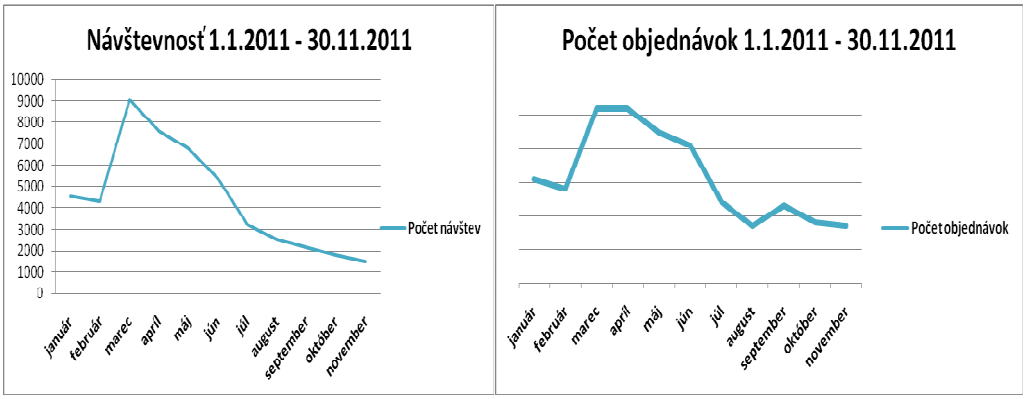
V tomto článku bude prezentovaný spôsob využitia vhodných metód internetového marketingu pre zvýšenie návštevnosti a obratu internetového obchodu firmy zaoberajúcej sa predajom a montážou satelitnej techniky a príslušenstva.

Táto firma už v minulosti identifikovala potenciál internetového predaja, preto v roku 2009 bol vytvorený ich prvý internetový obchod, ktorý však už v dnešnej dobe nevyhovoval súčasným trendom a jeho kvalita už takisto nebola postačujúca. Ročné náklady na toto riešenie boli vysoké, pričom dlhodobo neposkytovalo požadované výsledky, návštevnosť a obrat mali stále klesajúcu tendenciu. Tento fakt motivoval návrh projektu, v rámci ktorého by bol vytvorený prehľadný a dobre organizovaný internetový obchod, ktorý by oslovil široké spektrum zákazníkov. V rámci tohto projektu by bol zároveň navrhnutý efektívny marketingový mix pre danú firmu, ktorý by viedol k zvýšenie počtu zákazníkov a objednávok pri minimálnych nákladoch. Teda vynaložené finančné prostriedky na internetový marketing predstavovali dôležitý faktor projektu.

A. Pôvodná verzia obchodu

Podľa nižšie uvedených grafov (viď. Obr.1) je možné vidieť klesajúci trend návštevnosti a objednávok. Mesiace február, marec a apríl dosiahli najvyššie hodnoty – dôvod takejto vysokej návštevnosti a počtu objednávok pripisujeme vypínaniu analógového signálu a masovému prechodu na digitálny a satelitný príjem TV v tomto období.

[1] http://www.seomoz.org/beginners-guide-to-seo/growing-popularity-and-links
[2] http://www.uspesne.podnikam.sk/online-marketing/akych-podnikatelov-hodi-reklama/21264



Obr. 1 Návštevnosť a počet objednávok pôvodnej verzie internetového obchodu.

Spolu s majiteľmi firmy boli identifikované tieto základné problémy:

- zastaraný dizajn
- nedostatočná funkcionalita
- návštevnosť a obrat ma dlhodobo klesajúcu tendenciu
- nízke pozície vo vyhľadávačoch
- žiadna internetová reklama
- slabá propagácia na sociálnych sieťach a chýbajúci email marketing
- vysoká konkurencia v odvetví
- nízky počet objednávok
- vysoké náklady na prevádzku systému

B. Návrh novej verzie

Nový elektronický obchod, spĺňajúci väčšinu aktuálnych trendov a požiadaviek na správny a efektívny web, bolo možné implementovať pomocou dvoch variant: vytvorenie kompletného riešenia na mieru alebo použitie voľne dostupného systému. Po zvážení viacerých kritérií, časového hľadiska a požiadaviek firmy bola vybraná druhá možnosť. Do užšieho výberu sa dostali tieto systémy: PrestaShop, osCommerce, Drupal e-commerce, Zen-Cart, Magento, VirtueMart – Joomla, OpenCart. Všetky uvedené systémy spĺňali skoro všetky požiadavky, avšak ako finálne riešenie bol vybraný systém OpenCart[3], v tom čase vo verzii 1.5.0. Zaujala nás hlavne prehľadná a jednoduchá administrácia, široká ponuka modulov a veľká česko-slovenská užívateľská komunita.

III. INTERNETOVÝ MARKETING

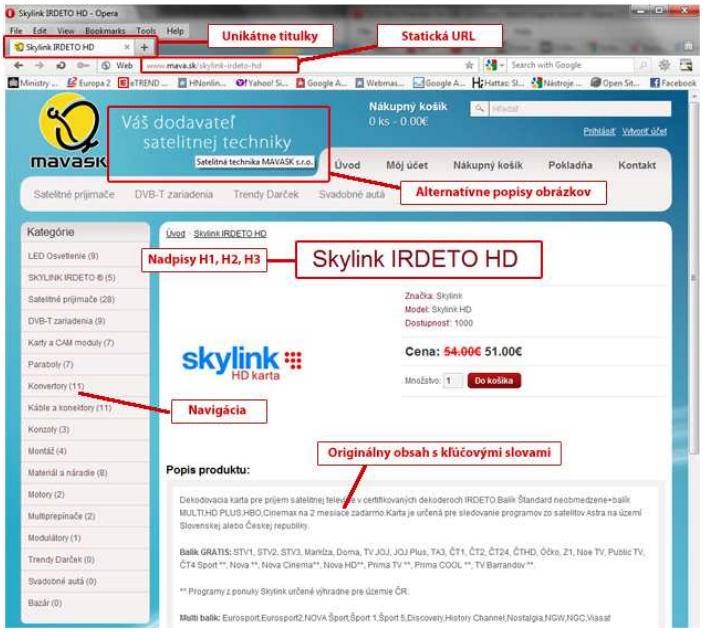
A. Implementácia Google Analytics

Po inštalácií systému, nasadení dizajnu a spustení obchodu do plnej prevádzky bol na systém nasadený analytický nástroj Google Analytics[4]. Tento systém je najpoužívanejší a najkomplexnejší nástroj na meranie a analyzovanie návštevnosti internetových stránok. Možnosti, ktoré ponúka sú mnohonásobne vyššie v porovnaní s inými meracími nástrojmi. Okrem analýzy návštevnosti nám ponúka komplexné údaje o návštevníkoch a ich správaní. Pomocou týchto údajov môžeme získať dôležité údaje na optimalizáciu jednotlivých podstránok, odhadnutie trendov vo vyhľadávaní alebo na monitoring SEO optimalizácie. Google Analytics je taktiež nevyhnutným nástrojom pri monitorovaní všetkých druhov on-line reklamy a ich efektívnosti.

B. Aplikácia SEO

Na obr. 2 môžeme vidieť základné techniky on-page optimalizácie internetového obchodu. Celá on-page optimalizácia bola zameraná na zvýšenie hustoty kľúčových slov a optimalizáciu titulku hlavnej stránky.

[3] http://www.opencart.com/
[4] http://www.google.com/analytics/



Obr. 2 On–page optimalizácia hlavnej stránky internetového obchodu.

Pôvodná verzia internetového obchodu mala málo spätných odkazov, z čoho plynul nielen nižší Pagerank (algoritmus hodnotenia webových stránok na základe ich liniek s inými stránkami) ale aj nižšie pozície vo vyhľadávaní. Prvým krokom off-page optimalizácie bol teda linkbuilding - tvorba spätných odkazov pomocou registrácie do slovenských katalógov stránok. Odkazy smerovali priamo na hlavnú stránku elektronického obchodu, aby táto stránka dosiahla čo najvyšší rank. Pre registráciu boli vytvorené 3 rôzne popisy stránky, aby sa obmedzila duplicita, čo znižuje hodnotu odkazu. Stránka bola zaregistrovaná spolu do 50 katalógov stránok a 6 vyhľadávačov tovarov. Okrem vyhľadávačov boli spätné odkazy vložené na partnerské webové stránky tejto firmy – napr. www.trendydarcek.sk, www.bombadarcek.sk a ďalšie. Pri tomto procese sa dbalo na anchor text odkazu a na optimalizáciu pomocou kľúčového slova „satelitná technika“.

C. Realizácia PPC reklamy

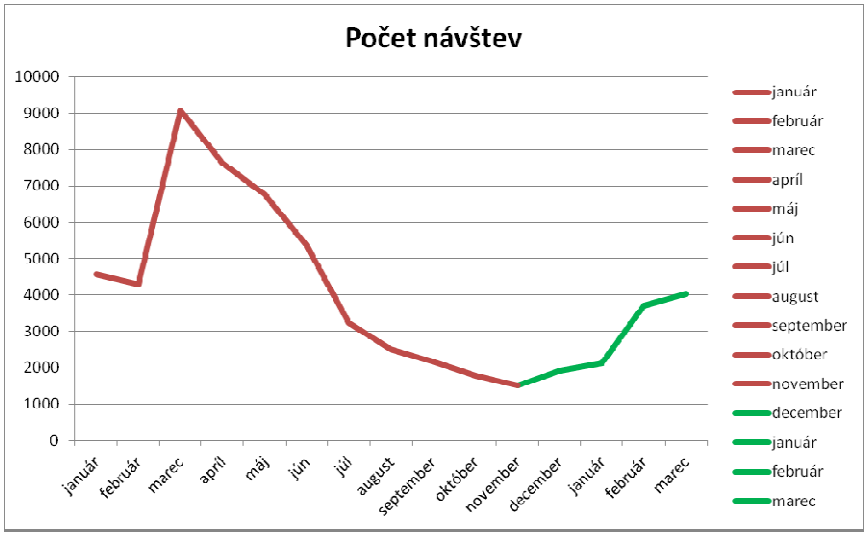
Pre tieto účely boli použité dve možnosti: Heureka.sk a Google AdWords. Prvá alternatíva bola zaujímavá z pohľadu, že podľa prieskumov tento portál navštevujú práve technicky zdatnejší užívatelia na vyhľadávanie elektroniky. Okrem vyhľadávania a porovnávania tovarov Heureka.sk ponúka PPC systém na prednostné vypisovanie obchodov vo výsledkoch vyhľadávania. Z pomedzi všetkých druhov propagácie internetových obchodov, PPC reklama vo vyhľadávačoch tovarov dosahuje najvyššiu mieru konverzie a ponúka najvyššiu mieru návratnosti investícií, preto bola pre tento obchod spustená testovacia kampaň s rozpočtom 30€.

V rámci Google Adwords boli vytvorené dve reklamné jednotky, ktoré sa zamerali na dve produktové skupiny „Satelitné príjimače“ a „LED Osvetlenie“. Pomocou nástroja pre návrh kľúčových slov od spoločnosti Google boli vybrané relevantné kľúčové slová, na ktoré sme sa následne zamerali vo vyhľadávači Google.

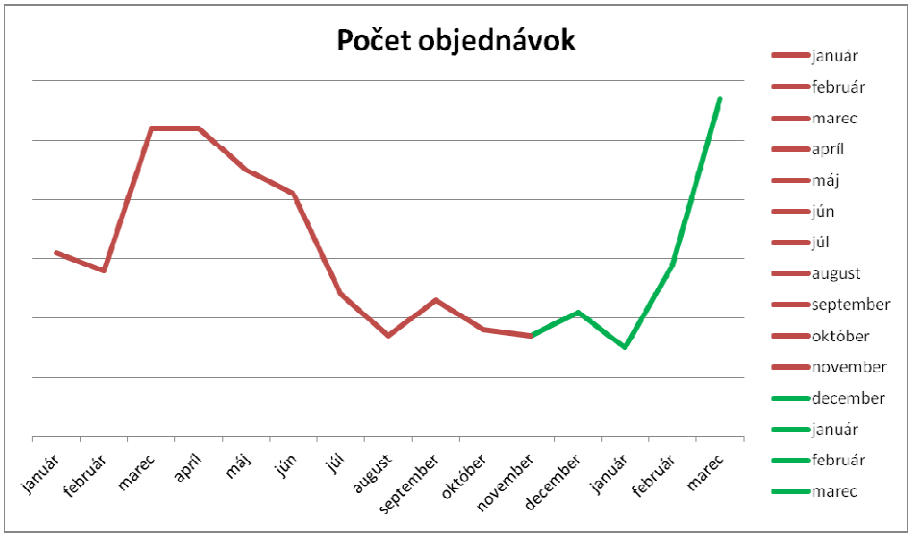
IV. ZÁVER

Navrhnutý a realizovaný mix optimalizačných techník a PPC reklamy dosiahol dobré výsledky aj pri nízkom rozpočte - 70€. Počet návštev sa vplyvom optimalizácie postupne začal zvyšovať, takýto trend je predpokladaný aj v najbližších mesiacoch, keďže vplyv optimalizácie sa prejavuje postupne.

Ako môžeme vidieť na nasledujúcich grafoch, efektivita predaja sa zvýšila o viac ako 100%. Pri návštevnosti približne 4000 užívateľov bolo v marci vytvorených viac objednávok ako v tom istom čase pred rokom pri návštevnosti približne 9000 užívateľov. Hlavným dôvodom je príchod relevantných zákazníkov – z vyhľadávačov tovarov a z vyhľadávača Google. Z toho môžeme usúdiť, že návštevnosť pri internetových projektoch nie je všetko a je potrebné zaoberať sa inými metrikami - miera konverzie, počet stránok na návštevu, miera odchodu zo vstupnej stránky, analýza výstupných stránok a efektivita reklamy.



Obr. 3 Zvýšenie návštevnosti vplyvom SEO optimalizácie a PPC reklamy.



Obr. 4 Zvýšenie počtu objednávok vplyvom vyhľadávačov tovarov a PPC reklamy.

V tabuľke č.1 môžeme vidieť zvýšenie pozícií vo vyhľadávači Google na vybrané kľúčové slová. Pri týchto kľúčových slovách je *„SEO konkurencia“* pomerne vysoká, preto pre dosiahnutie prvých pozícií, je potrebný dlhší čas a rozpočet pre reklamnú kampaň. Dôležité je stále pokračovať v optimalizácií stránky, keďže pri stagnácii nás môže konkurencia predbehnúť. SEO nie je momentálny stav, ale neustály proces.

Tabuľka 1
Zvýšenie pozícií vo vyhľadávačoch na vybrané kľúčové slová elektronického obchodu

Kľúčové slovo	6.12.2011	31.3.2012
satelitná technika prešov	14.	2.
antenna technika	24.	16.
satelitná technika	64.	5.
satelitna technika	78.	8.
satelitné prijímače	65.	22.
satelitné prijímače skylink	35.	9.

POĎAKOVANIE

Tento príspevok vznikol za podpory KEGA projektu č. 065TUKE-4/2011 (50%) a realizáciou projektu Rozvoj Centra informačných a komunikačných technológií pre znalostné systémy (kód ITMS projektu: 26220120030) na základe podpory operačného programu Výskum a vývoj financovaného z Európskeho fondu regionálneho rozvoja (50%).

LITERATÚRA

[1] Elektronické podnikanie, 1. vydanie, 2011, Dostupné na: http://www.scribd.com/doc/51220745/%C4%8CASOPIS
[2] Blog o eshopoch a internetovom predaji. Dostupné na: http://www.eshopar.sk/2011/06/vyhody-a-nevyhody-e-shopov-na-mieru/
[3] M. Kubíček, J. Linhart, *333 tipů a triků pro SEO*, Computer Press 2010, p.19.

Analýza informačného systému firmy na základe IT Scorecard

Mária BACHLEDOVÁ, Adela TUŠANOVÁ, Ján PARALIČ

Katedra kybernetiky a umelej inteligencie, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

maria.bachledova@tuke.sk, adelatusanova@tuke.sk, jan.paralic@tuke.sk

***Abstrakt* — IT Balanced Scorecard je metodika hodnotenia efektívnosti investícií do informačných systémov, resp. IT služieb. U nás na Slovensku metodika nie je veľmi známa, ale v posledných rokoch zaznamenáva celosvetovo široké využitie v rôznych firmách. Cieľom našej práce je analýza informačného systému konkrétnej firmy práve na základe metódy IT Scorecard. V prvej časti podávame úvod do problematiky metodík merania výkonnosti, a to Balanced Scorecard a IT Scorecard. V druhej časti analyzujeme firmu NAJAVO, s.r.o., pre ktorú bola navrhnutá koncepcia vo forme IT Scorecard. V tretej časti popisujeme navrhovanú á koncepciu pre konkrétny informačný systém firmy.**

***Kľúčové slová* — Balanced scorecard, IT scorecard**

I. BALANCED SCORECARD A IT SCORECARD

Balanced Scorecard (ďalej BSC) vytvorili v 90-tych rokoch Robert Kaplan a David Norton. Táto metóda vytvára väzbu medzi strategickými cieľmi firmy a jej operatívnymi činnosťami, cieľom objektívne merať ich výkon, a to nielen pomocou tradičných finančných ukazovateľov, ale v celom komplexe. Hlavnou úlohou BSC nie je hľadať a formulovať vhodnú stratégiu firmy, ale pomôcť jej k implementácii prostredníctvom definovania cieľov a meradiel v štyroch perspektívach: finančná perspektíva, perspektíva interných procesov, zákaznícka perspektíva a perspektíva učenia a rastu. Pre jednotlivé perspektívy je charakteristické vzájomné logické prepojenie a práve tento základ poskytuje komplexný pohľad na organizáciu a na jej výkonnosť. Zdôrazňovaná je tiež skutočnosť, že tradičné meradlá minulej výkonnosti sú doplnené o nové meradlá budúcej výkonnosti. BSC je preto chápaný ako vyvážený systém (balanced), pretože udržiava rovnováhu medzi finančnou a nefinančnou oblasťou. [1] [3]

Pri tvorbe a implementácii Balanced Scorecard podnik vychádza zo zvolenej stratégie. Na základe stratégie stanoví podnik strategické ciele v štyroch základných perspektívach a priradí k nim meradlá úspešnosti ich plnenia (kvantifikovateľné ukazovatele). Následne si podnik stanoví cieľové hodnoty pre jednotlivé meradlá a definuje akčné plány na ich dosiahnutie.

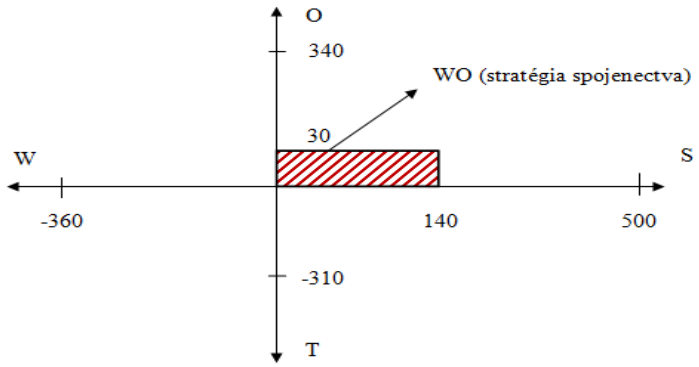
V posledných rokoch sa stale viac na meranie výkonnosti informačných systémov v podniku používa metóda IT Balanced Scorecard, ktorej základ vychádza zo systému merania výkonnosti metódy BSC. IT Scorecard je systém merania výkonnosti, ktorý sa zameriava na informačné systémy a technológie (IT). IT Scorecard tvoria taktiež 4 perspektívy, ktoré boli pretransformované z pôvodnej koncepcie BSC na tieto štyri perspektívy: celopodnikový prínos, orientácia na zákazníka, prevádzková dokonalosť a orientácia na budúcnosť. [2][4][5]

V tejto práci sme sa snažili vytvoriť optimálny návrh vo forme IT Scorecard, ktorý by pomohol firme zhodnocovať aj nefinančné ukazovatele a maximalizovať úžitok z vybraného informačného system s použitím vhodných merateľných ukazovateľov špecifických pre danú firmu.

II. ANALÝZA FIRMY NAJAVO S.R.O.

Firma NAJAVO, s.r.o. pôsobí na trhu v oblasti informačných technológií. Vznikla v roku 1990 a počas svojej existencie vyvinula programy pre jednoduché a podvojné účtovníctvo, ekonomicko-informačný systém Modular Business Manager či Najavo Professional. Okrem toho vyvinula taktiež svoj vlastný podnikový informačný systém na skladové hospodárstvo NAJAVO, ktorý poskytuje informácie o účtovníctve a sklade. Práve tento informačný systém v práci ďalej analyzujeme pomocou metódy IT scorecard.

Pre tvorbu logickej koncepcie vo forme IT Scorecard bolo najskôr nevyhnutné vykonať analýzu firmy, nakoľko IT scorecard vychádza z vízie a stratégie firmy ako celku. Dôležitá je tiež aktuálna situácia firmy, preto sme urobili SWOT analýzu, ktorej výsledok uvádzame na Obr. 1.



Obr. 1 SWOT analýza firmy NAJAVO, s.r.o.

Analýzami sme zistili, že firma sa nachádza v I. kvadrante - v stratégií spojenectva, t.j. slabé stránky prevládajú nad silnými a príležitosti prevládajú nad hrozbami. Odporučili by sme využiť príležitosti na odstránenie slabých stránok podniku, pretože podnik sa nachádza v konkurenčnom prostredí. Je potrebné posilniť taktiež svoju aktuálnu pozíciu na trhu a odstrániť nedostatky.

Nosnou stratégiou firmy je výroba a predaj hlavných výrobkov a služieb s cieľom maximalizovať spokojnosť konečných zákazníkov. Firma sa kontinuálne usiluje smerovať k naplneniu stratégie, zvyšovať svoju výkonnosť, zefektívniť procesy vo firme a získať lepšiu pozíciu na trhu.

Cieľom spoločnosti je tiež rozvíjať nosný produkt Najavo Professional, reagovať na požiadavky zákazníkov ako aj, na technologické zmeny a poskytovať také produkty, ktoré umožnia zákazníkom získať konkurenčnú výhodu na trhu. Pri dosahovaní tohto cieľa je veľmi nápomocná metóda IT Scorecard, ktorej návrh pre informačný systém NAJAVO uvádzame v kapitole III.

III. NÁVRH IT SCORECARD PRE INFORMAČNÝ SYSTÉM NAJAVO

Prvotná časť práce predstavovala prevod stratégie na čiastkové strategické ciele podľa jednotlivých perspektív metódy. Vypracovaný návrh bol následne prevedený do grafickej formy, tzv. strategickej mapy. Pritom bola dodržaná forma príčina - následok. Posledná časť práce predstavovala kompletný pohľad na zvolenú implementovanú metodiku pri jednotlivých perspektívach, vrátane merateľných ukazovateľov a firemných cieľov pre ich hodnoty.

K implementácií metódy bola dôležitá stratégia podniku, ktorá bola účastníkmi prevedená na strategické ciele a vytvoril sa kompletný návrh vo forme IT Scorecard s cieľom zhodnocovania podnikového IS pomocou finančných a nefinančných ukazovateľov a s cieľom dosiahnutiu rovnováhy (balance) v podniku.

Firma dospela k strategickým cieľom prostredníctvom otázok, ktoré sme pre nich pripravili a spresnením vízie spoločnosti, ktorých výsledkom boli strategické ciele zobrazené prostredníctvom strategickej mapy (nástroj pre grafické znázornenie spojitostí strategických cieľov), viď Obr. 2.



Obr. 2 Strategická mapa firmy NAJAVO, s.r.o. pre IS NAJAVO

V nasledujúcej časti budú popísané strategické ciele podľa jednotlivých perspektív definovaných metódou IT Scorecard. Každá perspektíva obsahuje tri strategické ciele s presným popisom strategických akcií. Návrh bol firmou schválený.

A. Celopodnikový prínos

Na Obr. 3 je znázornená tvorba IT Scorecard z pohľadu prvej perspektívy – celopodnikový prínos. Spôsob ako znížiť náklady na IS firmy spočíval v obmedzení používania technických zariadení, v odstraňovaní nadbytočných procesov a vo zvýšení tržieb z predaja výrobkov a služieb. Firma ako celok potrebuje získať zákazníkov cez výrazný marketing, preto zvolenou alternatívou sa stala alternatíva formou propagácie produktov cez efektívny guerilla marketing, čo predstavuje efektívnu reklamu za nízke náklady. [6]

Strategický cieľ	Merateľný ukazovateľ	Popis meradla	Cieľová hodnota a merné jednotky
Celopodnikový prínos			
Zníženie nákladov na IS	Celkové náklady na IS	Ukazovateľ vyhodnocovaný ročne na základe porovnania nákladov s predchádzajúcim obdobím	5 % ročne (EUR, %)
Zvýšenie rastu tržieb na IS	Celkové tržby na IS	Ukazovateľ vyhodnocovaný ročne na základe porovnania výsledkov s predchádzajúcim obdobím	10 % ročne (EUR, %)
Zvýšenie rentability tržieb z IS	Rentabilita tržieb na IS	Ukazovateľ vyhodnocovaný ročne na základe porovnávania hodnôt za jednotlivé roky	10 centov ročne (EUR,%)

Obr. 3 Tvorba IT Scorecard pre IS NAJAVO

B. Orientácia na zákazníka

Jediný spôsob ako zvyšovať spokojnosť zákazníka je realizácia pravidelného prieskumu spokojnosti zákazníka s cieľom odstraňovania skutočností, s ktorými zákazníci nie sú spokojní. Získavanie nových zákazníkov je možné zavedením vhodnej propagácie produktov a vyžaduje si preskúmanie trhu a nájdenie takých cieľových zákazníkov, pre ktorých bude firmou pripravený marketingový mix efektívny. Firma poskytuje taktiež dva typy programov, ktoré sú najžiadanejšie z pohľadu zákazníkov. Každoročne vznikajú na základe požiadaviek zákazníkov

nové moduly pri jednotlivých programoch, s ktorými vznikajú aj chyby, ktoré je potrebné odstrániť. Cieľom je pozorovať úspešnosť splnenia požiadaviek zákazníkov. Návrh na 90 % úspešnosť plnenia požiadaviek spočíva v pravidelnom rokovaní so zákazníkmi o presných požiadavkách a uskutočňovanie pravidelných stretnutí s cieľom zistenia aktuálneho stavu žiadaného programu respektíve konzultácií o potrebných zmenách v programe. Nasledujúci obrázok (viď Obr. 4) predstavuje IT Scorecard z pohľadu druhej perspektívy – orientácia na zákazníka.

Strategický cieľ	Merateľný ukazovateľ	Popis meradla	Cieľová hodnota a merné jednotky
Orientácia na zákazníka			
Zvýšenie spokojnosti zákazníkov	Dotazník	Dotazník vyhodnocovaný raz ročne	85 % spokojnosť zákazníkov
Získanie nových zákazníkov	Počet uzavretých zmlúv	Ukazovateľ vyhodnocovaný ročne	1 nový zákazník do 1 roka
Špičkové riešenia podľa požiadaviek zákazníkov	Úspešnosť splnenia požiadaviek zákazníka	Ukazovateľ vyhodnocovaný ročným počtom vzniknutých a odstránených chýb pri jednotlivých programoch	90 % úspešnosť plnenia požiadaviek zákazníkov

Obr. 4 Tvorba IT Scorecard pre firmu NAJAVO

C. Prevádzková dokonalosť

Koncepcia pre tretiu perspektívu a jej podklady pre tvorbu strategických akcií je zobrazená na Obr. 5.

Strategický cieľ	Merateľný ukazovateľ	Popis meradla	Cieľová hodnota a merná jednotka
Prevádzková dokonalosť			
Zvýšenie výkonnosti odberateľských procesov	Úspešnosť nasadenia zákazky vo firme	Ukazovateľ vyhodnocovaný pomerom štandardnej dĺžky nasadenia programu ku skutočnej dĺžke nasadenia programu vyjadrený v %	90 % úspešnosť zamestnancov (%)
Skrátenie doby cyklu dodávky pri špeciálnych zákazkách	Doba od uzavretia zmluvy so zákazníkom po dodanie zákazky zákazníkovi	Ukazovateľ vyhodnocovaný rozdielom dátumu inštalácie programu a dátumu uzavretia zmluvy	4 dní odovzdanie každej zákazky skôr (dni)
Zásadný rast prostredníctvom inovácií	Počet záujemcov o nové moduly	Ukazovateľ vyhodnocovaný ročne podľa počtu záujemcov o moduly s cieľom rozširovania modulov s najväčším počtom záujemcov	2 záujemcovia ročne (počet)

Obr. 5 Tvorba IT Scorecard pre firmu NAJAVO

Všetky zákazky, ktoré firma implementuje pri svojich zákazníkoch trvajú určitú dobu. Každý program má štandardnú dĺžku, ako dlho sa implementuje program do spustenia jeho ostrej prevádzky. Tým, že sa bude porovnávať skutočná dĺžka implementácie programov so štandardnou dĺžkou vyjadrená v %, bude možné ľahko sledovať výkonnosť firmy v tomto ukazovateli. Čím rýchlejšie bude program implementovaný vo firme a bude implementovaný správne, tým výkonnejší bude odberateľský proces.

Návrh skrátenia doby cyklu dodávky pri špeciálnych zákazkách spočíva v pravidelnom kontrolovaní úspešnosti podľa požiadaviek zákazníka s cieľom kvalitného a včasného spracovania úlohy.

Inovácie predstavujú dôležitý aspekt udržania sa firmy na trhu. Firma každoročne vyvinie podľa požiadaviek zákazníka okolo 6 modulov, pričom o každý modul je určitý počet záujemcov. Návrh spočíva v rozvíjaní tých modulov, o ktoré majú zákazníci najväčší záujem a ktorý prinesie firme pozitívne finančné výsledky.

D. Orientácia na budúcnosť

Poslednou perspektívou, ktorej prislúcha návrh IT Scorecard je zobrazený na Obr. 6. Skutočnosťou, že firma bude zameriavať časť svojho úsilia na zvyšovanie výkonnosti zamestnancov, zaistí nie len zníženie nákladov, ale zefektívnia sa tiež procesy vo firme. Všetky procesy, ktoré sa vo firme uskutočňujú vychádzajú z kvality zamestnancov. Návrh spočíva v pozorovaní a vyhodnocovaní počtu vypracovaných dotazov testermi z celkového počtu ročných dotazov, a vypracovaných dotazov programátormi, vyjadrené v %. Výkonnostná úroveň podniku vychádza z kvalifikácie a schopnosti zamestnancov prispôsobovať sa a úrovne ich schopností. Tým, že firma bude poskytovať školenia zamestnancom, bude zvyšovať ich kvalifikáciu, ktorá je predpokladom úspešnosti firmy. Taktiež prieskum o motivovaní a vedení zamestnancov k efektívnym pracovným výkonom umožňuje prispôsobovať sa a plniť potreby svojich zamestnancov, ktorí musia byť motivovaní k tomu, aby požadovaný výkon dosiahli. Navrhnutý prieskum bude realizovaný jeden až dvakrát ročne s cieľom zistenia spokojnosti a motivácie zamestnancov.

Strategický cieľ	**Merateľný ukazovateľ**	**Popis meradla**	**Cieľová hodnota a merná jednotka**
Orientácia na budúcnosť			
Zvýšenie výkonnosti zamestnancov	Percentuálne vyjadrenie výkonnosti zamestnancov	Ukazovateľ vyhodnocovaný ročne evidovaním počtov zákazníckych dotazov, počtom vypracovaných dotazov testermi a programátormi vyjadrené v %	90 % výkonnosť zamestnancov, (%)
Zvýšenie výkonnosti podniku na základe kvalifikácie zamestnancov	Ročný počet školení na zamestnanca	Ukazovateľ vyhodnocovaný ročne	1-2 školenia ročne na zamestnanca, (počet)
Efektívne riadenie ľudských zdrojov	Dotazník	Ukazovateľ vyhodnocovaný na základe prieskumu motivácie zamestnancov	95% spokojnosť a motivácia zamestnancov

Obr. 6 Tvorba IT Scorecard pre firmu NAJAVO

IV. IMPLEMENTÁCIA NAVRHNUTÉHO IT SCORECARD PRE INFORMAČNÝ SYSTÉM NAJAVO

Po návrhu IT Scorecard bolo potrebné navrhnúť a implementovať vhodnú softvérovú podporu pre čo najjednoduchiu realizáciu pravidelného hodnotenia informačného systému NAJAVO vo firme. Do úvahy pripadali viaceré existujúce riešenia, pričom sme analyzovali štyri hlavné možnosti, a síce komerčný softvér BSC Designer, aplikácia pomocou Google Docs, aplikácia pomocou MS Excel a vlastná webová aplikácia. Vzájomné porovnanie týchto štyroch alternatívnych implementácií prezentuje Obr. 7.

Po zasadnutí s riaditeľom firmy a po posúdení jednotlivých kritérií účastníci dospeli k názoru, že optimálne riešenie pre firmu bude vytvorenie aplikácie v Microsoft Excel. Zakúpenie softvéru BSC Designer firma odmietla z dôvodu neprimeraných finančných nákladov. Web aplikácia a nástroj Google Docs bola pre firmu zaujímavá, ale odmietla ju z dôvodu závislosti od internetového pripojenia a kvôli bezpečnosti interných dát. Poslednou alternatívou bolo vytvorenie aplikácie v Microsoft Excel, s ktorou riaditeľ firmy súhlasil. Dôvodom súhlasu boli nasledujúce skutočnosti - nepotrebné internetové pripojenie, istá bezpečnosť dát, spoločnosť už má zakúpený licencovaný MS Office, praktické skúsenosti zamestnancov firmy s MS Excel.

Pripadajúca alternatíva	Kritéria				
	Cena	Školenie	Funkcie	Internetové pripojenie	Bezpečnosť dát
Web aplikácia	Podľa zložitosti aplikácie	áno	Tvorba ukazovateľov Merné jednotky Vykresľovanie grafov	potrebné	neistá
Legálny softvér BSC Designer	648 €	388 €	Strategická mapa Tvorba ukazovateľov Merné jednotky Vykresľovanie grafov	nepotrebné	istá
Aplikácia pomocou MS Excel	Zdarma	Zdarma	Tvorba ukazovateľov Vykresľovanie grafov Upozornenia pri prekročení negatívnych hodnôt	nepotrebné	istá
Aplikácia pomocou Google Docs	zdarma	zdarma	Tvorba dokumentov Tvorba tabuliek Tvorba prezentácií Kresieb a formulárov	potrebné	neistá

Obr. 7 Výber optimálneho riešenia pre firmu NAJAVO na základe zvolených kritérií

V. ZÁVER

Cieľom tohto článku bolo predstaviť návrh spôsobu hodnotenia vybraného informačného systému pre firmu NAJAVO, s.r.o. s cieľom poukázať na skutočnosť možnej implementácie metódy formou IT Scorecard. Základ práce bolo prevedenie stratégie podniku na čiastkové strategické ciele v jednotlivých perspektívach a podanie návrhu strategických akcií s cieľom zefektívnenia procesov vo firme a nájdenie optimálneho riešenia pre implementáciu metódy. Ako najvhodnejšie riešenie softvérovej podpory bol vybraný MS Excel, v ktorom bola spracovaná metóda implementovaná.

POĎAKOVANIE

Táto práca bola podporovaná Vedeckou grantovou agentúrou Ministerstva školstva, vedy, výskumu a športu Slovenskej republiky a Slovenskej akadémie vied na základe zmluvy č. 1/1147/12 (50%) a realizáciou projektu Rozvoj Centra informačných a komunikačných technológií pre znalostné systémy (ITMS kód: 26220120030) na základe podpory operačného programu Výskum a vývoj financovaného z Európskeho fondu regionálneho rozvoja (50%).

LITERATÚRA

[1] BOBENIČ HINTOŠOVÁ, A.: Nový pohľad na balanced scorecard, Manažment v teórii a praxi, on-line odborný časopis o nových trendoch v manažmente, ISSN 1336-7137.
[2] Keyes, J.: Implementing the IT Balanced Scorecard: Aligning IT with Corporate Strategy, ISBN-10: 0849326214
[3] Kaplan R S and Norton D P (1992): The Balanced Scorecard: measures that drive performance, Harvard Business Review, Vol. 70, Nr. 1 (Jan – Feb), pp. 71–80
[4] Van Grembergen, Wim - Saull, Roland. IT BSC at a Major Canadian Financial Group. [Online] http://www.performance-measurement.net/news-detail.asp?nID=35.
[5] Van Grembergen W. and De Haes S., 2005, Measuring and improving Information Technology Governance through the Balanced Scorecard, Information Systems Control Journal, vol. 2
[6] LEVINSON, J. C., LAUTENSLAGER, A. LAUTENSLAGER. Guerilla marketing in 30 days. Irvine, CA: Entrepreneur Media, 2005. 267 s. ISBN 1-932531-29-7.

Využitie umelej inteligencie v spracovaní obrazu

Matúš BAČO

Katedra kybernetiky a umelej inteligencie, Fakulta elektrotechniky a informatiky,
Technická univerzita v Košiciach, Slovenská republika

matus.baco@student.tuke.sk

***Abstrakt* — Cieľom práce bolo navrhnúť, naprogramovať a v praxi overiť systém seba-lokalizácie pre potreby robotického futbalu, ktorého základ tvorí neurónová sieť so spätným šírením chyby. Dáta pre neurónovú sieť budeme získavať s pomocou algoritmov počítačového videnia, ktoré obsahuje určité algoritmy aj na predspracovanie obrazu. Experimentmi by sme mali overiť najvhodnejšie parametre pre danú neurónovú sieť a porovnať ich. Výsledky sú zobrazené pomocou grafov a pravdepodobnosti úspešnosti pre jednotlivo vykonávané experimenty.**

***Kľúčové slová* — robot Nao, počítačové videnie, neurónové siete, seba-lokalizácia, backpropagation**

I. Úvod

Spoľahlivá seba-lokalizácia je hlavná požiadavka na autonómne mobilné roboty. Za posledných pár rokov bolo vynájdených mnoho algoritmov a techník na splnenie tejto úlohy. Robocup je zaujímavá vedecká výzva pre študentov po celom svete. Doménou Robocupu je skúšať nové algoritmy a demonštrovať existujúce techniky v prostredí robofutbalu. V Robocupe, spolupráca robotov vedie k inteligentnému skupinovému správaniu, takýto typ správania zvyšuje šancu vyhrať zápas v robotickom futbale. Spolupráca spočíva v komunikácii medzi jednotlivými robotmi, výmenou informácií ako sú šance na skórovanie, taktické rozmiestňovanie, poloha brány, pohyb protivníkov, poloha lopty atď. Rámcová referencia alebo poloha robota vo štvorci, vytvára úlohu nájsť polohu robota relatívne k ihrisku. Táto úloha je neľahká, nakoľko robot sa nachádza v dynamickom prostredí, počas zápasu sa vyskytuje pomerne dosť rušivých vplyvov, a len máloktoré z nich vieme eliminovať. V našom laboratóriu CIT v Košiciach, máme robota Nao, ktorý je používaný ako štandard, v robotickej lige Standard Platform League. Na základe predchádzajúcich poznatkov som sa rozhodol vytvoriť systém seba-lokalizácie pre robota Nao, ktorý v prvej fáze vytvorí model prostredia, pomocou ktorého sa dokáže robot v druhej fáze neskôr lokalizovať na ihrisku. Väčšina bežných metód využíva segmentáciu objektov vo vstupnom obraze. Vstupný obraz býva často zašumený. Na odhalenie pozície robota budeme používať nami vytvorený model sveta, ktorý bude uložený v synaptických prepojeniach neurónovej siete.

II. Analýza riešenia

Jedna z hlavných výziev súťaže RoboCup je, že do roku 2050 humanoidné roboty porazia obyčajných ľudí vo futbale. Popri tejto výzve okrem RoboCupu pomaly vyrastajú rôzne iné hry ako Robo Rescue alebo Robo Dance. Futbalová liga je rozdelená do niekoľkých líg, v ktorej každá má vlastný súbor pravidiel. Jedna z týchto líg je aj (SPL) Standard Platform League, kde každý tím má rovnaký model robota. Roboti sú plne autonómny, a nemajú žiadne externé ovládanie. Momentálne ako štandard v SPL je používaný robot Nao od spoločnosti Aldebaran.

V Robocupe sú vlastnosti ako rozmer ihriska, farba brán alebo čiar dopredu určené. Vďaka tomu dokážeme vytvoriť lokalizáciu s vopred známou mapou. Pohybom po ihrisku , robot pozoruje črty. Črty sú špecifické body ktoré identifikujú špecifickú polohu na ihrisku. Napríklad: Žltá tyč je videná v určitej výške len keď som v blízkosti žltej brány. Takmer vo všetkých algoritmoch seba-lokalizácie, sú črty popísané vzdialenosťou a uhlom k danej črte, získané z otáčania hlavy. Keďže namerané dáta bývajú často krát skreslené (nesprávna klasifikácia čŕt alebo nesprávny uhol alebo vzdialenosť) potom sa seba-lokalizácia sa stáva náročnejšou. Seba-lokalizácia v SPL je o to zaujímavejšia pretože tímy majú k dispozícii

rovnaký hardware, ktorý zahŕňa aj obmedzené zdroje ako je pamäť a výpočtová sila.

Automatická identifikácia objektov vo videu je významná oblasť výskumu v robotike a jej hlavnou časťou je senzorické vnímanie sveta.

Základné požiadavky na systém seba-lokalizácie pre potreby SPL

Systém spracovania videnia musí byť dostatočne rýchly, aby dosahoval čo najlepšie dáta pre ostatné časti systému. To znamená, že by mal byť schopný spracovať viac ako 30 snímkou za sekundu.

Mal by byť spustiteľný na Nao-vom 500Mhz procesore.

Objekty by mali byť identifikované s dostatočnou presnosťou

Mal by byť značne robustný voči vysokému rozmazaniu.

Musí eliminovať detekciu falošných objektov

Mal by byť dostatočne odolný voči čiastočne zakrytým objektom

Systém na identifikáciu objektov je zvyčajne založený na klasifikátore farieb spoločne s detektorom rohov. Klasifikácia farieb umožňuje rýchlu identifikáciu približnej oblasti objektu, ktorá je veľmi presná v obraze, a poskytuje značnú robustnosť na nerovnomerne sa meniace svetelné podmienky.

Tabuľka 1 Porovnanie jednotlivých tímov RoboCupu a ich metód

Použitá metóda	**Výhody**	**Nevýhody**	**Tímy ktoré používajú metódu**
(rozšírená) MCL	Preukázala sa ako presná, poradí si z odpočinkovým problémom , poradí si z komplexným odhadom.	Výpočtovo drahá	Austrian-Kangaroos, B-human, Cerebrus, Edinferno, Noxious-Kouretes, TJArk
MCL & Mosr	Lepšie výsledky ako MCL/sr		Cmurfs
MCL & neg. Inform.	Rýchlejšia eliminácia častíc ako MCL		TT-UT Austin Villa
MCL & KF	Nižšia výpočtová cena ako MCL		rUNSWift
AUX PF & SIR			SPQR+Uchile
vzdialenosť od bránkových tyčí	Jednoduchosť	Nepresné	L3M, NTU RobotPal
UKF & MH	Funguje hladko a dobre		Nao Devils Dortmund
viacnásobný RKF	výpočtovo nenáročný		SPIteam
Lokalizácia obmedzení	Nízka výpočtová cena, lepší výkon ako častícový filter		Nao Team Humboldt
Rao-Black & KF	Nízka výpočtová cena, rýchlejšia znovu-lokalizácia		Upennalizers
systém s lokálne citlivým správaním	Jednoduchosť	Ťažkosti zo spoľahlivosťou	Wrighteagle Unleashed!
Lokálny Model	Jednoduchosť	žiadna komunikácia medzi robotmi	WPI Warriors
Cox & CI & UKF	vysoký potenciál	nestabilné zatiaľ	RoboEireann

III. NEURÓNOVÁ SIEŤ

Pre úlohu klasifikácie sú neurónové siete veľmi silným nástrojom, pretože plnia funkciu univerzálneho aproximátora. Aby sme mohli využívať neurónovú sieť je dobré vedieť ako neurónová sieť funguje. Neurónová sieť je všeobecný paralelný procesor, ktorý má tendenciu uchovávať poznatky, a ďalej ich vie využívať. Neurónové siete napodobňujú ľudský mozog v 2 aspektoch:

Poznatky sú získavané počas učenia

Medzineurónové spojenia (synaptické váhy) sú používané na ukladanie znalostí

Inšpirácia pri tvorbe neurónových sietí sa čerpala s ľudského mozgu, a to v princípe činnosti a štruktúre prepojení. Teória opisujúca neurónové siete vychádza z neurofyziologických poznatkov. Vysvetľuje základný princíp spracovania informácie v nervových bunkách. Niekedy sa neurónové siete označujú aj ako model mozgu bez mysle, keďže sa snažia pochopiť nervový systém, ale nezaoberajú sa psychikou.

IV. MODEL NEURÓNU

Modely neurónu sú väčšinou abstrakciou mechanizmu, ako nervové bunky spracovávajú informácie. Nie je možné vytvoriť presný postup, „matematicky popísať“ schopnosti reálneho neurónu a to z viacerých dôvodov. Výpočtové procesy na neurónoch sú enormne zložité, obsahujú asynchrónne procesy a dva základne odlišné postupy na prenosu vzruchu: neurotransmisiou (krátkodobé ovplyvnenie postsynaptického neurónu) a neuromoduláciou (dlhodobé ovplyvňovanie postsynaptického neurónu). Jeden neurón môže obsahovať až niekoľko desiatok 10^4 prepojení . Pričom význam týchto váh zvyšuje ich lokalizácia. Synapsie uložené bližšie k iniciálnemu segment majú väčší vplyv na dosiahnutie excitácie ako vzdialenejšie. Pri súčasnej aktivácii dvoch susedných synaptických prepojení, sa tvoria na membránach komplexné málo preskúmané fenomény(“skraty”). Väčšina týchto procesov má nelineárny priebeh. Veľkosť siete je z výpočtového hľadiska veľmi významný parameter. Komplexnosť formálneho neurónu komplikuje aj matematickú analýzu správania sa siete. Fyziológia neurónu je celkom nepreskúmaná oblasť. Doposiaľ nie je možné jednoznačne opísať funkciu neurónu ako výpočtového elementu. Na obrázku je základný model biologického neurónu.

V. EXPERIMENTY

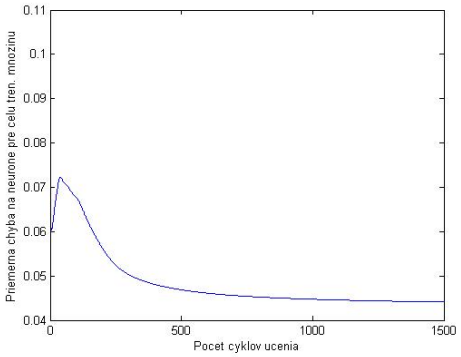
Návrh experimentov pozostáva z experimentálneho hľadania najlepšej topológie siete. Počas experimentov sa budú meniť parametre ako vstupný počet neurónov, počet neurónov na skrytých vrstvách, parameter učenia a strmosť sigmoidálnej funkcie. Experiment budeme vykonávať na vzorke 100 obrázkov, 5 rôznych obrázkov pre každú pozíciu. Túto databázu rozdelíme v pomere 3:2 , pre každú pozíciu, prvá skupina bude tvoriť súbor dát pre získanie modelov siete. Dáta je potrebne generovať pri každej zmene vstupných parametrov. Pre testovaciu množinu použijeme dáta s 2 obrázkov pre každú pozíciu.

Experiment č. 1

Vstupnú vrstvu tvorí mriežka o veľkosti 6x8 neurónov.

Parametere experimentu:

- Strmosť sigmoidy = 1
- Parameter učenia gamma = 0.05
- Počet neurónov na 1. Skrytej vrstve 48
- Počet neurónov na 2. Skrytej vrstve 48
- Počet cyklov učenia 1500



Obr. 1 Chyba učenia pre celú trénovaciu množinu experiment č. 1.

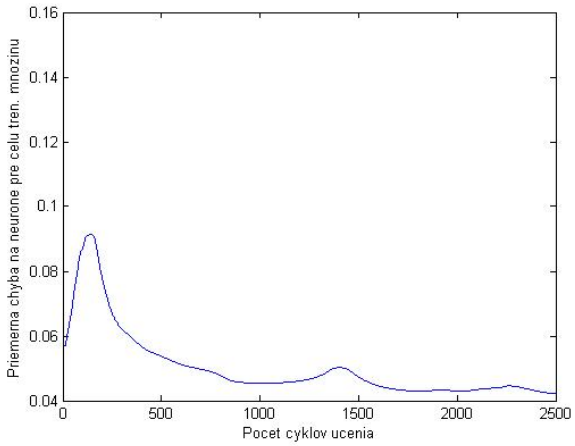
Priemerná úspešnosť výslednej klasifikácie pre experiment č.1 : **30%**

Experiment č. 2

Vstupnú vrstvu tvorí mriežka o veľkosti 6x8 neurónov.

Parametre experimentu:

- Strmosť sigmoidy = 1
- Parameter učenia gamma = 0.01
- Počet neurónov na 1. Skrytej vrstve 24
- Počet neurónov na 2. Skrytej vrstve 24
- Počet cyklov učenia 2500



Obr. 2 Chyba učenia pre celú trénovaciu množinu experiment č. 2.

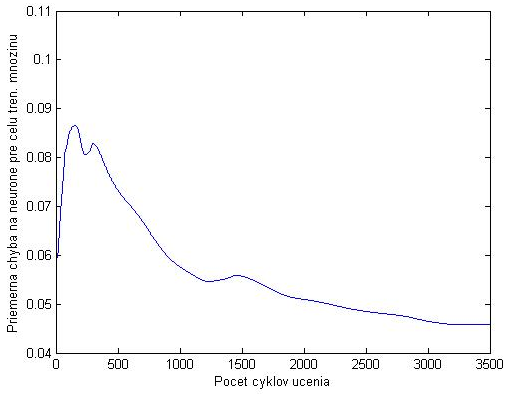
Priemerná úspešnosť výslednej klasifikácie pre experiment č.2 : 4**0%**

Experiment č.3

Vstupnú vrstvu tvorí mriežka o veľkosti 6x8 neurónov.

Parametere experimentu:

- Strmosť sigmoidy = 1
- Parameter učenia gamma = 0.01
- Počet neurónov na 1. Skrytej vrstve 24
- Počet neurónov na 2. Skrytej vrstve 24
- Počet cyklov učenia 3500



Obr. 3 Chyba učenia pre celú trénovaciu množinu experiment č. 3

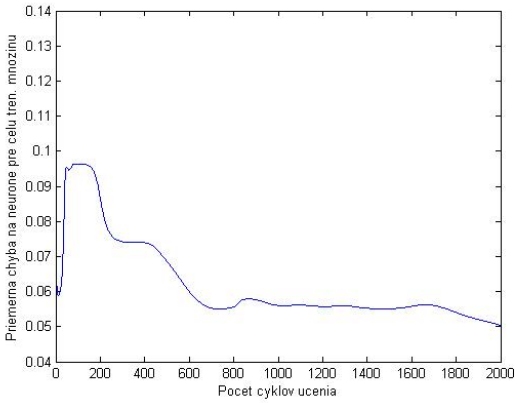
Priemerná úspešnosť výslednej klasifikácie pre experiment č.3 : 4**0%**

Experiment č.4

Vstupnú vrstvu tvorí mriežka o veľkosti 6x8 neurónov.

Parametere experimentu:

- Strmosť sigmoidy = 0.5
- Parameter učenia gamma = 0.1
- Počet neurónov na 1. Skrytej vrstve 24
- Počet neurónov na 2. Skrytej vrstve 24
- Počet cyklov učenia 2000



Obr. 4 Chyba učenia pre celú trénovaniu množinu experiment č. 4.

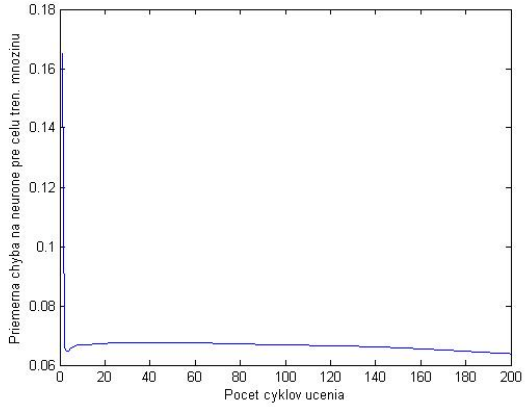
Priemerná úspešnosť výslednej klasifikácie pre experiment č.4 : 5**0%**

Experiment č. 5

Vstupnú vrstvu tvorí mriežka o veľkosti 12x16 neurónov.

Parametre experimentu:

- Strmosť sigmoidy = 0.5
- Parameter učenia gamma = 0.05
- Počet neurónov na 1. Skrytej vrstve 100
- Počet neurónov na 2. Skrytej vrstve 100
- Počet cyklov učenia 200



Obr. 5 Chyba učenia pre celú trénovaciu množinu experiment č. 5

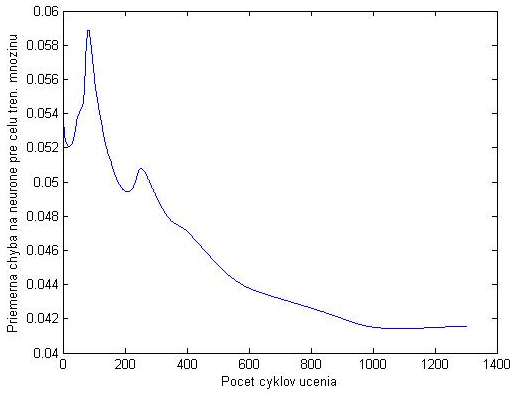
Priemerná úspešnosť výslednej klasifikácie pre experiment č.5 : **35%**

Experiment č. 6

Vstupnú vrstvu tvorí mriežka o veľkosti 12x16 neurónov.

Parametre experimentu:

- Strmosť sigmoidy = 1
- Parameter učenia gamma = 0.2
- Počet neurónov na 1. Skrytej vrstve 70
- Počet neurónov na 2. Skrytej vrstve 70
- Počet cyklov učenia 1300



Obr. 6 Chyba učenia pre celú trénovaciu množinu experiment č. 6.

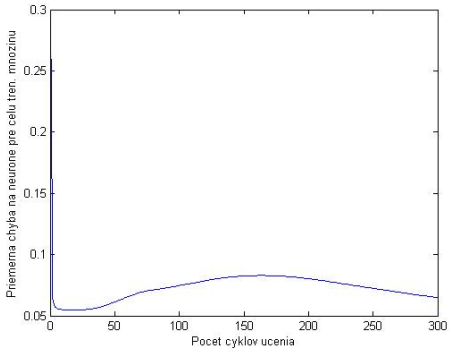
Priemerná úspešnosť výslednej klasifikácie pre experiment č.6 : **40%**

Experiment č. 7

Vstupnú vrstvu tvorí mriežka o veľkosti 12x16 neurónov.

Parametere experimentu:

- Strmosť sigmoidy = 0.5
- Parameter učenia gamma = 0.05
- Počet neurónov na 1. Skrytej vrstve 50
- Počet neurónov na 2. Skrytej vrstve 50
- Počet cyklov učenia 300



Obr. 7 Chyba učenia pre celú trénovaciu množinu experiment č. 7

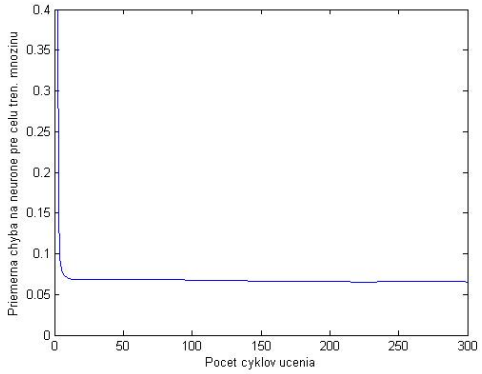
Priemerná úspešnosť výslednej klasifikácie pre experiment č.6 : **35%**

Experiment č. 8

Vstupnú vrstvu tvorí mriežka o veľkosti 12x16 neurónov.

Parametere experimentu:

- Strmosť sigmoidy = 1
- Parameter učenia gamma = 0.01
- Počet neurónov na 1. Skrytej vrstve 50
- Počet neurónov na 2. Skrytej vrstve 50
- Počet cyklov učenia 520



Obr. 8 Chyba učenia pre celú trénovaciu množinu experiment č. 8.

Priemerná úspešnosť výslednej klasifikácie pre experiment č.8 : **35%**

VI. ZÁVER

Nakoľko sme vykonali celú radu experimentov, dospeli sme k záveru že systém seba-lokalizácie pomocou neurónových sietí nie je vhodný pre použitie v SPL, nakoľko nespĺňa základnú požiadavku a to tú aby bol systém autonómny. Počas experimentov sme tiež zistili, že algoritmus je neefektívny, pretože najvyššia spoľahlivosť dosahovala len 50% a v tomto prípade je systém nepoužiteľný. Zistené výsledky boli čiastočne ovplyvňované selekciou prahovej hodnoty, čiastočne aj pozadím v laboratóriu CIT kde sme vykonávali experimenty ako aj náhodným výberom obrázkov z databázy. Lepšie výsledky by sa dali dosiahnuť zvýšením vstupných čŕt do neurónovej siete, napríklad pridaním dodatočných informácií, výška tyčí a dĺžka brvna. Ako skvelý doplnok by bolo dobré vytvoriť aj algoritmus ktorý určí aj polohu lopty na ihrisku. Časové hľadisko experimentov nebudeme hodnotiť nakoľko čas potrebný na vykonanie experimentu sa líšil v závislosti od počtu cyklov a úmerne aj od celkového počtu neurónov. Úspešnejšie výsledky by sme mohli dosiahnúť rozšírením systému o ďaľšiu neurónovú sieť, ktorá by spracovávala dáta, extrahované napríklad z čiar. Tieto dve neurónové siete by mali potom možnosť hlasovať za aktuálnu pozíciu na ihrisku. Alternatívou k nášmu systému by som navrhoval implementovať metódu Monte Carlo, nakoľko túto metódu používa

momentálne väčšina tímov v SPL. Niektoré tímy využívajú len lokálnu seba-lokalizáciu, ktorá je vo väčšine prípadov postačujúca.

POĎAKOVANIE

Táto práca bola vytvorená realizáciou projektu Rozvoj Centra informačných a komunikačných technológií pre znalostné systémy (kód ITMS projektu: 26220120030) na základe podpory operačného programu Výskum a vývoj financovaného z Európskeho fondu regionálneho rozvoja. Výskum podporovaný Národným projektovým grantom pre výskum a vývoj 1/0667/12 „Inkrementálne metódy učenia pre inteligentné systémy“ 2012-2015.

LITERATÚRA

[1] Austrian-Kangoroos tímový dokument, Team description paper Austrian-Kangaroos, Vienna University of Technology and University of Applied Sciences Technikum Vienna. Dostupné na internete: <https://docs.google.com/viewer?a=v&pid=sites&srcid=YXVzdHJpYW4ta2FuZ2Fyb29zLmNvbXxwdWJsaWNfZW58Z3g6MjQxM2YzZTFkYjE4OWNjNg>
[2] [2]P.L. ROSIN a G.A.W. WEST,. Nonparametric segmentation of curves into
[3] various representations. Pattern Analysis and Machine Intelligence, IEEE
[4] Transactions on, 17(12):1140-1153, dec 1995.
[5] [3] S. THRUN, W. BRGARD, a D. FOX. Probabilistic robotics. Intelligent
[6] robotics and autonomous agents. MIT Press, 2005.
[7] [4] Thomas RÖFER, Tim LAUE, Colin GRAF, Tobias KASTNER, Alexander FABISCH,Christian THEDIECK, Team description for Robocup 2010, Deutsches Forschungszentrum f ür Künstliche Intelligenz, Bremen. Dostupné na internete:
[8] < http://www.b-human.de/downloads/bhuman10_tdp.pdf>
[9] [5] J.-S. GUTMANN a D. FOX. An experimental comparison of localization
[10] methods continued. In Intelligent Robots and Systems, 2002. IEEE/RSJ
[11] International Conference on, volume 1, pages 454 - 459 vol.1, 2002.
[12] [6] M. MONTEMERLO, S. THRUN, D. KOLLER, and B. WEGBREIT. FASTSLAM: A factored solution to the simultaneous localization and mapping problem. In Proceedings of the National conference on Articial Intelligence, pages 593-598. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press;1999, 2002.
[13] [7] M. QUINLAN and R. MIDDLETON. Multiple model kalman lters: A localization technique for robocup soccer. RoboCup 2009: Robot Soccer World Cup XIII, strany 276-287, 2010.
[14] [8] NTU RoboPAL tímový dokument, Chieh-Chih Wang, Shao-Chen Wang, Chun-Hua Chang, Bo-Wei Wang, Hsin-Cheng Chao and Chih-Chung Chou, Team Report and Code Release 2011, Graduate Institute of Networking and Multimedia
[15] National Taiwan University, Taipei, Taiwan. Dostupná na internete:
[16] <http://www.csie.ntu.edu.tw/~bobwang/RoboCupSPL/NTURoboPAL_TeamReport2011.pdf>
[17] [9]S. MCDONALD J. WHELAN, T. Stüdli and R. MIDDLETON. Line point registra-
[18] tion: A technique for enhancing robot localization in a soccer environment.
[19] In Proceedings RoboCup Symposium, strany 2538-2543, 2011.
[20] [10] rUNSWift tímový document, Adrian Ratter, Bernhard Hengst, Brad Hall, Brock White, Benjamin Vance, Claude Sammut, David Claridge, Hung Nguyen, Jayen Ashar,
[21] Maurice Pagnucco, Stuart Robinson, Yanjin Zhu, rUNSWift Team Report 2010
[22] Robocup Standard Platform League, University of New South Wales , Australia
[23] Dostupná na internete:
[24] < http://www.cse.unsw.edu.au/~robocup/2010site/reports/report2010.pdf>
[25] [11]SPIteam tímový dokument, U.Rey Juan Carlos, U. Leon, U. Rovira i Virgili,
[26] SPITeam’s Application for qualification to RoboCup-2011, University Tarragona, Dostupná na internete:
[27] < http://www.spiteam.org/documents/SPIteam-qualification-2011.pdf>
[28] [12]SPQR-UChile tímový dokument, P. Guerrero, J. M. Yañez, R. Palma-Amestoy, R. Marchant, S. Blunier, J.Ruiz-del-Solar, RoboCup 2011 Standard Platform League Team Description Paper, Universidad de Chile, Dostupná na internete:
[29] < http://www.tzi.de/spl/pub/Website/Teams2011/UChile-SPQR_tdp2011.pdf>
[30] [13] TJArk tímový dokument, The Application of TJArk for Standard Platform
[31] League in 2011, Tongji University, Shanghai 201804, P.R.China, Dostupná na internete:
[32] <http://see.tongji.edu.cn/TJArk/download/papers/TJArkNaoApp2011.pdf>
[33] [14] HESSEL VAN DER MOLEN, Self-Localization in the RoboCup Soccer Standard Platform League with the use of Dynamic Tree, University Of Amsterdam 2011
[34] < http://staff.science.uva.nl/~bredeweg/pdf/BSc/20102011/VanDerMolen.pdf>
[35] [15]Upennalizers tímový dokument, Jordan Brindza, Levi Cai, Ashleigh Thomas,
[36] Ross Boczar, Alyin Caliskan, Alexandra Lee, Anirudha Majumdar, Roman Shor, Barry, Scharfman, Dan Lee RoboCup Standard Platform League Team Report 2010, Universityof Pennsylvania. Dostupná na internete:
[37] <http://www.seas.upenn.edu/~robocup/files/upennalizers_team_research_report_2010.pdf>
[38] [16] KRÁĽ, Andrej-TIŇO,Peter-FARKAŠ, Igor-POSPÍCHAL, Jiří-BEŇUŠKOVÁ, Ľubica-KVASNIČKA, Vladimír: Úvod do teórie neurónových sietí, 1997, Dostupná na internete:
[39] <http://ics.upjs.sk/~novotnyr/home/skola/neuronove_siete/nn_kvasnicka/Uvod%20do%20NS.pdf >
[40] [17] SINČÁK, Peter – ANDREJKOVÁ, Gabriela : Neurónové siete : Inžiniersky
[41] prístup (1.diel). Košice : TU-FEI, 1996. 117 s. Dostupná na internete:
[42] <http://www.ai-cit.sk/source/publications/books/NS1/html/index.html>
[43] [18] RATH Christof, Self-localization of a biped robot in the RoboCup Standard Platform League domain, Graz University of Technologi

Manažment sémantických znalostí

Lucia BRÉDOVÁ, Marián MACH

Katedra kybernetiky a umelej inteligencie, Fakulta elektrotechniky a informatiky,
Technická univerzita v Košiciach, Slovenská republika

lucia.bredova@student.tuke.sk, marian.mach@tuke.sk

***Abstrakt* — Článok prezentuje návrh riešenia pre interaktívny zber údajov v štruktúre riadenej sémantickým modelom v oblasti prípravy simulačných modelov multiagentových systémov. Podľa návrhu je implementovaný funkčný prototyp, ktorý umožňuje užívateľovi zobrazovať informácie o načítanej ontológii, vytvárať, mazať, upravovať inštancie, a to všetko zjednodušeným spôsobom pomocou odvodených znalostí získaných prostredníctvom reasonera RacerPro. Funkčnosť prototypu je otestovaná na vytvorenej ontológii vo formáte OWL v doméne sociálnych simulácií.**

***Kľúčové slová* — OWL ontológia, RacerPro, Jracer, OWLAPI**

I. Úvod

Jedným z hlavných problémov webu je, že drvivá väčšina informácií na ňom zverejnených je navrhovaná len pre človeka, nie je však zrozumiteľná pre stroje. Z tohto dôvodu bol vytvorený koncept sémantického webu. Táto hierarchicky koncipovaná vrstvová architektúra [1] smeruje k spravovaniu, manipulácii a využívaniu informácií na logickej a sémantickej úrovni. Vytvorenie sémantických znalostí v konkrétnej oblasti predpokladá vytvorenie formalizovaného konceptuálneho modelu určitého výseku reality - ontológie.

Cieľom práce je návrh a implementácia OWL editora inštancií, ktorý ponúka užívateľovi funkcie ako sú načítanie ontológie, zobrazenie informácií o triedach, vlastnostiach a inštanciách danej ontológie, vytváranie, mazanie a úprava inštancií, a to zjednodušeným spôsobom pomocou odvodených znalostí a nakoniec export vytvorených údajov do RDF/XML alebo OWL/XML súboru. Vytvorený prototyp bol otestovaný na vytvorenej ontológii.

II. Sémantické znalosti

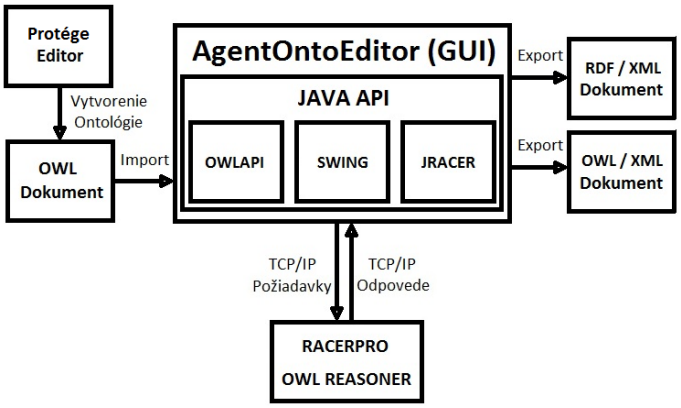
Vytváraním tried, definovaním vzťahov medzi nimi a opisom vlastností prvkov týchto tried je možné (napr. v jazyku OWL) modelovať určitý systém znalostí o vybranej oblasti záujmu. Kým triedy a indivíduá vytvárajú akúsi štruktúru ontológie, vlastnosti vnášajú do celej koncepcie isté tvrdenia o triedach a ich prvkoch. Vlastnosti predstavujú v OWL binárne relácie buď medzi dvoma objektmi alebo medzi objektom a hodnotou údajového typu. Vlastnosti môžu byť navzájom inverzné, funkcionálne, inverzne funkcionálne, tranzitívne a symetrické.

K ohraničeniu inštancií, ktoré patria k danej triede sa používajú reštrikcie vlastností. V OWL spadá ohraničenie do troch nasledovných kategórií: ohraničenie kvantifikátorom, kardinálne ohraničenie a ohraničenie *hasValue* (má hodnotu). Ohraničenie kvantifikátorom môže byť existenčné alebo všeobecné. Kardinálne ohraničenie sa používa k určeniu počtu relácií, v ktorých môže participovať inštancia s danou vlastnosťou. Kardinálne ohraničenia obsahujú nasledovné typy ohraničení: minimálne kardinálne ohraničenie, maximálne kardinálne ohraničenie a kardinálne ohraničenie. [2]

Medzi štandardnú funkcionalitu existujúcich riešení patrí dopytovanie sa nad metadátami a pre tento účel bolo vyvinutých niekoľko dopytovacích jazykov. Dôležitou súčasťou existujúcich riešení je taktiež odvodzovanie znalostí získaných pomocou reasonera (odvodzovača). V praktickej časti práce je používaný reasoner RacerPro a ten ponúka dopytovací jazyk nRQL [3].

III. Návrh riešenia

Cieľom bolo navrhnúť a implementovať systém pre interaktívny zber údajov v štruktúre riadenej sémantickým modelom pre cieľovú oblasť prípravy simulačných modelov multiagentových systémov. Na nasledujúcom obrázku je znázornená architektúra implementovaného systému:



Obr. 1 Architektúra navrhovaného svstému

V editore Protége sa vytvorí ontológia v doméne sociálnych simulácií vo formáte OWL. Tento OWL dokument sa následne načíta do aplikácie AgentOntoEditor (AOE), kde si užívateľ bude môcť zobraziť informácie o ontológii a čo je hlavným cieľom vytváranej aplikácie, bude môcť vytvárať, mazať a upravovať inštancie ontológie čo najjednoduchším spôsobom prostredníctvom odvodených znalostí pomocou reasonera RacerPro. Nakoniec bude mať užívateľ možnosť vyexportovať ontológiu do RDF/XML alebo OWL/XML dokumentu.

Na realizáciu aplikácie je potrebné zvoliť komponenty, ktoré umožnia správne fungovanie systému:

- výber OWL editora potrebného na vytvorenie ontológie (Protége)
- výber implementačnej technológie (napr. J2EE (Platforma Java 2, Enterprise Edition))
- výber OWL reasonera (RACERPro)
- výber knižnice, vhodnej pre prácu s OWL údajmi (napr. OWLAPI)

Samotná realizácia aplikácie pozostáva z častí:

- vytvorenie ontológie vo formáte OWL pre cieľovú oblasť prípravy simulačných modelov multiagentových systémov
- vytvorenie samotnej aplikácie realizujúcej zber údajov
- vyhodnotenie vytvoreného prototypu na konkrétnej úlohe.

IV. Implementácia riešenia

A. Vybudovanie ontológie

Prvou úlohou pri implementácii riešenia je vybudovanie ontológie vo formáte OWL pre cieľovú oblasť prípravy simulačných modelov multiagentových systémov. Tomuto kroku predchádza analýza domény sociálnych simulácií.

Multiagentový systém je heterogénna množina agentov, medzi ktorými sú definované konkrétne vzťahy a väzby. V tomto systéme má každý agent určité prostriedky, s ktorými môže disponovať a zároveň aj obmedzenia, v ktorých hraniciach sa musí pohybovať pri dosiahnutí cieľa. Multiagentové sociálne simulácie, anglicky „Agent Based Social Simulation (ABSS)“ sa zaoberajú simuláciami sociálnych fenoménov prostredníctvom počítačov. Človek je modelovaný a vystupuje ako agent v softvérovom prostredí. ABSS využívajú multiagentové simulácie, počítačové simulácie a výsledky zo sociálnych vied. V kontexte sociálnych simulácií predstavujú agenti zjednodušené modely určitých úloh (hľadanie potravy, vzájomná komunikácia, predaj a nákup komodít, ap.). Vďaka tomuto zjednodušeniu často dokážu odhaliť určitú vlastnosť alebo typ správania systému, ktorú môžeme neskôr zovšeobecniť aj pre skutočný, ucelený a komplexný systém. [4]

Pri vytváraní ontológie bol využitý editor Protége. V prvom rade bolo potrebné vytvoriť triedy a ich vlastnosti. Medzi hlavné triedy patria triedy *Agent (agent), Property (vlastnosť), Object (objekt), Relation (vzťah), Facet (aspekt), Event (udalosť), Action (akcia), Message (správa), Rule (pravidlo), Condition (podmienka), Situation (situácia), Function (funkcia).*

Agent má podtriedy *AbstractAgent (abstraktný agent)* a *IndividualAgent (individuálny agent)*. Každý agent je definovaný nejakou vlastnosťou. Definovaná vlastnosť je založená na atribúte (binárny, symbolický alebo enumeratívny), je limitovaná obmedzením (enumeratívnym alebo typovým), ktoré dovoľuje určité hodnoty alebo typ a určená je práve jednou z týchto hodnôt. Objekt je tiež definovaný nejakou vlastnosťou.

Každá udalosť, ktorá nastane, je zapríčinená nejakou akciou, a tie môžu zmeniť vlastnosť agenta alebo objektu. Udalosť sa môže prihodiť nejakému objektu, na ktorom je vykonávaná akcia, ktorá túto udalosť zapríčinila. Akcia môže byť vykonávaná agentom, takisto ho môže aj ovplyvňovať. Agent môže prijímať i odosielať správy, ktoré sú dodávané pravidlami. Pravidlá sú vyvolané nejakou situáciou, ktorá pozostáva z podmienok.

B. Samotná realizácia aplikácie

Na import a export OWL dokumentu je použitá knižnica OWLAPI, ktorá ponúka metódy na realizáciu týchto operácií. Takisto ponúka metódy i na manipuláciu s údajmi obsiahnutými v načítanej ontológii.

Na získanie informácií o načítanej ontológii je použitý reasoner RacerPro prostredníctvom knižnice JRacer, ktorému je poslaná požiadavka, ontológia je následne klasifikovaná a reasoner pošle späť odpoveď na dopyt vo forme stringového reťazca, ktorý je pre správne zobrazenie nutné rozparsovať. Na dopytovanie ponúka reasoner RacerPro dopytovací jazyk nRQL. Pomocou komponentov knižnice SWING je výsledok graficky zobrazený užívateľovi v ľahko čitateľnej forme, či už v podobe rozbaľovacích stromov, zoznamov, popisiek či textových polí.

V. TESTOVANIE A VYHODNOTENIE VYTVORENÉHO PROTOTYPU

Za demonštračný príklad vybraný na vyhodnotenie vytvoreného prototypu je použitá simulácia známej rozprávky o troch prasiatkach. Na tomto príklade je možné jednoduchým spôsobom prezentovať vlastnosti vytvorenej ontológie.

Hlavnými inštanciami, ktoré sú vytvárané pomocou editora AgentOntoEditor sú zobrazené v nasledujúcej tabuľke:

Tabuľka 1 Zoznam hlavných tried a inštancií

Triedy	AbstractAgent	IndividualAgent	Property	Object
Inštancie	Wolf	Wolf1	Location	House
	Pig	MotherPig	Alive	Material
	LittlePig	Pig1	PigAttitude	
		Pig2	Substance	
		Pig3		

Ďalšími dôležitými inštanciami sú hodnoty, ktoré môžu jednotlivé vlastnosti nadobúdať:

Tabuľka 2 Inštancie triedy Value

Inštancie Property	Location	Alive	PigAttitude	Substance
Inštancie Value	PigHome	Yes	Mild	Straw
	World	No	Relaxed	Sticks
	WolfHome		Serious	Bricks
	LocStraw			
	LocSticks			
	LocBricks			

V neposlednom rade to sú inštancie tried *Action*, *Event* a *Message*, ktoré menia podmienky a stav prostredia.

Dôležitou súčasťou procesu zberu inštancií je odvodzovanie znalostí získaných pomocou reasonera RacerPro. Tieto odvodené znalosti zjednodušujú jednotlivé kroky procesu a uľahčujú tak prácu s editorom. Odvodzovanie znalostí je využívané hlavne pri vytváraní nových tvrdení, kedy je zoznam povolených tried obmedzený reštrikciami *allValuesFrom* a zoznam povolených indivíduí vyselektovaný prostredníctvom nRQL dopytov:

```
racer.racerAnswerQuery("(?Value)","(and(#!:" + currentSelectionField + " ?Constraint #!:limitedBy) (?Constraint ?Value #!:allows))");
```

Tento konkrétny dopyt je použitý pri vytváraní novej relácie *Property – assignedBy – Value (vlastnosť – je určená - hodnotou),* kedy sa inštancii triedy *Property* s vlastnosťou *assignedBy* priradí inštancia triedy *Value*. *CurrentSelectionField* je aktuálne vybraná inštancia triedy *Property* a symbol „*#!*“ nahrádza menný priestor.

V dialógovom okne, ktoré obsahuje zoznam povolených inštancií triedy *Value* by sa bez tohto dopytu zobrazili všetky hodnoty triedy *Value*. Cieľom použitia dopytu však je vymedziť len tie hodnoty, ktoré spĺňajú dané podmienky:

- *currentSelectionField – limitedBy – ?Constraint,* čiže aktuálne vybraná hodnota triedy *Property* je limitovaná nejakým obmedzením
- *?Constraint – allows - ?Value,* obmedzenie povoľuje určité hodnoty

Majme príklad s triedami a ich inštanciami, ktoré sú zobrazené v nasledujúcej tabuľke:

Tabuľka 3 Príklad dopytovania a odvodzovania

TRIEDY	PROPERTY	CONSTRAINT	VALUE
INŠTANCIE	WOLF_LOCATION	CONSTRAINT_ LOCATION_WOLF	WOLF_HOME
			PIG_HOME
			LOC_STRAW
			LOC_STICKS
			LOC_BRICKS
			WORLD

Ak inštancia WOLF_LOCATION je limitovaná inštanciou CONSTRAINT_LOCATION_WOLF, a tá povoľuje inštancie WOLF_HOME, LOC_STRAW, LOC_STICKS, LOC_BRICKS a ak sa do premennej *currentSelectionField* vloží inštancia WOLF_LOCATION, tak výsledkom dopytu budú hodnoty: WOLF_HOME, LOC_STRAW, LOC_STICKS, LOC_BRICKS. Vo výsledku budú chýbať hodnoty PIG_HOME a WORLD.

Na základe tohto dopytu a jemu podobných sú v aplikácii uľahčené viaceré kroky editovania, pretože užívateľ nemusí prehľadávať celý zoznam inštancií prislúchajúcich danému rozsahu vlastností, ale je mu ponúknutý kratší a tým pádom prehľadnejší zoznam.

Vytvorená ontológia je v závere procesu editácie vyexportovaná do OWL/XML súboru.

VI. Záver

Práca sa zaoberá oblasťou sémanticky reprezentovaných znalostí. V prvej časti bola predstavená táto problematika s dôrazom na OWL technológie.

Hlavným cieľom práce bolo vytvoriť systém pre interaktívny zber údajov v štruktúre riadenej sémantickým modelom pre cieľovú oblasť prípravy simulačných modelov multiagentových systémov. Tento cieľ bol zrealizovaný v podobe aplikácie AgentOntoEditor, ktorá ponúka užívateľovi funkcie ako sú načítanie ontológie, zobrazenie informácií o triedach, vlastnostiach a inštanciách ontológie, vytváranie, mazanie a úprava inštancií, zjednodušená editácia inštancií pomocou odvodených znalostí a export vytvorených údajov do RDF/XML alebo OWL/XML súboru.

Vytvorený systém bol otestovaný v rámci jeho použitia pre prípravu simulačných modelov v doméne sociálnych simulácií. Vďaka odvodeným znalostiam získaným prostredníctvom reasonera RacerPro bol tento proces tvorby a editácie modelu zjednodušený.

Poďakovanie

Táto práca bola vytvorená realizáciou projektu Rozvoj Centra informačných a komunikačných technológií pre znalostné systémy (kód ITMS projektu 26220120030) na základe podpory operačného program Výskum a vývoj financovaného z Európskeho fondu regionálneho rozvoja.

Literatúra

[1] Lucia Brédová, Ukážková aplikácia RDF technológií, Bakalárska práca, Košice, 2010, 56s.
[2] Opis jazyka OWL pre reprezentáciu ontológií. [cit. 2012-05-08]. Dostupné na internete: <http://www2.fiit.stuba.sk/~kapustik/ZS/Clanky0607/klempa/index.html#_Toc150072831>.
[3] nRQL Tutorial. [cit. 2012-05-08]. Dostupné na internete: <http://www.michael-wessel.info/papers/nRQL-tut3.pdf>.
[4] Multiagentní sociální simulace. [cit. 2012-05-08]. Dostupné na internete: <http://cs.wikipedia.org/wiki/Multiagentn%C3%AD_soci%C3%A1ln%C3%AD_simulace>.

Návrh využitia metodológie METHONTOLOGY pre aplikáciu vybraného konceptuálneho modelu

Peter BUTKA

Centrum pre informačné technológie, Katedra kybernetiky a umelej inteligencie, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská Republika

peter.butka@tuke.sk

***Abstrakt* — Táto práca sa zaoberá využitím konkrétnej metodológie budovania ontológie pre aplikáciu a znovupoužitie vybraného hybridného konceptuálneho modelu na báze formálnej konceptovej analýzy a zhlukovania. Vybraný konceptuálny model bol primárne určený pre vytvorenie/návrh systému vyhľadávania informácií s podporu znalostného modelu získaného metódami dolovania znalostí z textových dokumentov. Popis tohto modelu z pohľadu metodológie pre budovanie ontológie, kde vybranou bola konkrétne metodológia METHONTOLOGY, umožňuje jeho znovupoužitie v novej doméne alebo aplikácii.**

***Kľúčové slová* — metodológie tvorby ontológií, konceptuálne modely, vyhľadávanie informácií**

I. ÚVOD

V praxi existuje viacero prístupov k analýze množín textových dokumentov s cieľom vytvoriť konceptuálny model rozširujúci tradičný vektorový model množiny dokumentov o určité triedy (koncepty) a ich vzťahy. Aplikačný potenciál takýchto modelov je najlepšie viditeľný pri ich použití v procese vyhľadávania informácií. Jeden takýto hybridný konceptuálny model na báze FCA (Formal Concept Analysis – Formálna konceptová analýza, [4]) v kombinácii so zhlukovaním je stručne predstavený aj v tejto práci. Vytvorený model je ekvivalentom konceptuálneho modelu na ontologickej báze, ktorý je aplikovaný v procese vyhľadávania dokumentov (IR). Avšak hľadisko ďalšieho použitia navrhnutých metód, modelov a postupov si vyžaduje stanoviť aj určitý metodologický pohľad na možnosti realizácie takejto aplikácie v novej doméne či pre inú množinu dokumentov.

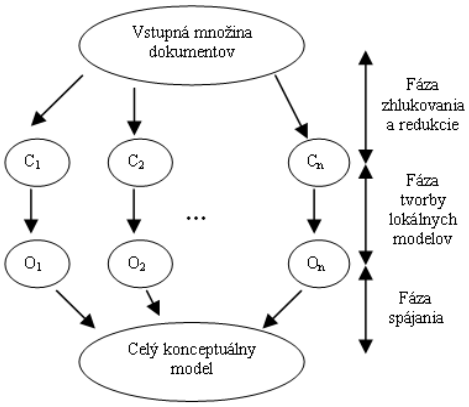
Vybranou metodológiou, ktorá bola aplikovaná a rozšírená pre naše potreby, je metodológia známa pod menom METHONTOLOGY [5]. V jednotlivých podkapitolách sú popísané práve spomenuté aspekty problému, t.j. stručný popis prvkov hybridného konceptuálneho modelu na báze FCA a zhlukovania, návrh systému pre jeho využitie v rámci IR, ako aj metodológia pre znovupoužitie prístupu v novej doméne, či pre novú množinu dokumentov.

II. POPIS HYBRIDNÉHO MODELU DOKUMENTOV NA BÁZE FCA A ZHLUKOVANIA

V našom ponímaní je možné využitie FCA pri tvorbe hierarchie konceptov a relácií medzi nimi. Veľkosť problému v doméne textových dokumentov nás privádza k myšlienke redukcie na subproblémy. Jednou z možností je zakomponovanie redukcie pomocou rozdeľujúcej procedúry zhlukovania. Potom by riešenie budovania konceptuálneho modelu mohlo pozostávať z nasledujúcich krokov (viď. Obr.1):

- Zhlukovacia fáza – predzhlukovanie vstupnej množiny dokumentov, nájdenie zaujímavých podskupín podobných dokumentov.
- Redukčná fáza – rozdelenie do podmnožín podľa predchádzajúceho kroku, príprava týchto podmnožín na ďalšie spracovanie, redukcia atribútov na základe rozdelenia do zhlukov (t.j. filtrácia termov pre jednotlivé zhluky podľa objektov v nich).
- Fáza budovania lokálnych modelov – každý zhluk je nezávislou trénovacou množinou, pre každý z nich sa vytvorí lokálny konceptuálny model.

- Fáza spájania – množina lokálnych modelov sa v tomto kroku spája do jedného veľkého modelu, vzťahujúceho sa k celej testovanej množine.

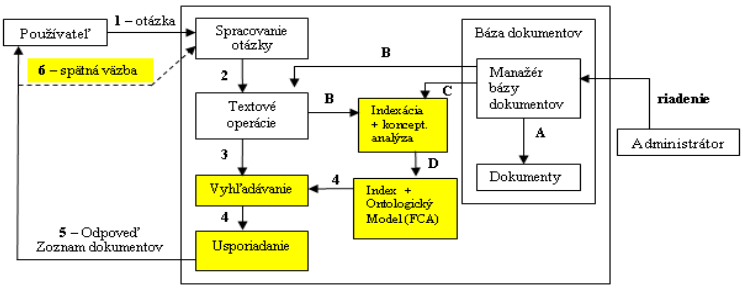


Obr. 1 Všeobecná schéma procesu tvorby konceptuálneho modelu dekompozíciou.

Základom konkrétnej realizácie konceptuálneho modelu je teda dekompozícia množiny dokumentov na menšie množiny použitím zhlukovacieho algoritmu (konkrétne bol použitý tzv. GHSOM). Následne sú pre každý zhluk vybudované pomocou FCA lokálne modely, ktoré sú kombinované do jednoduchej hierarchie konceptov použitím aglomeratívneho zhlukovania na základe podobnosti ich popisov. Pre budovanie lokálnych FCA modelov bol použitý algoritmus generovania horných susedov (upravený pre generovanie jednostranne fuzzy konceptového zväzu), tieto boli popísané charakteristickými termami a jednoduchý aglomeratívny zhlukovací algoritmus bol použitý pre budovanie hierarchie konceptových zväzov pomocou metriky založenej na týchto charakteristických popisoch. Podrobný popis celého procesu budovania modelu s príslušnými referenciami je možné nájsť v [2]. Základné prvky výsledného modelu teda sú lokálne konceptové zväzy pre jednotlivé podmnožiny dokumentov (získané zhlukovaním pomocou GHSOM-u), popisy lokálnych konceptových zväzov pomocou kľúčových slov a hierarchická štruktúra popisov zhlukovaných príslušným aglomeratívnym algoritmom. Tento model bol následne aplikovaný pre vytvorenie systému vyhľadávania informácií.

III. SYSTÉM VYHĽADÁVANIA INFORMÁCIÍ NA BÁZE VYBRANÉHO KONCEPTUÁLNEHO MODELU

Na Obr.2 je znázornená upravená schéma systému vyhľadávania informácií s vyznačenými zmenami pre použitie navrhnutého konceptuálneho modelu.



Obr. 2 Upravená schéma vyhľadávania informácií podľa potreby hybridného konceptuálneho modelu.

Schému je v zásade možné rozdeliť na dve základné časti – riadiacu časť a proces vyhľadávania. V prvom prípade sú zainteresované najmä bloky prepojené s riadiacou (administratívnou) funkčnosťou, pre ktoré sú použité veľké písmená A – D. V druhom prípade ide o proces samotného vyhľadávania, kde sa vytvorený index používa a generujú sa výstupy, odpovede na používateľské otázky. Postupnosť procesu je daná označením prepojení zainteresovaných blokov (číselne, 1 až 6).

Zmenami v tejto schéme vzhľadom k využitiu nového modelu prechádzajú bloky Vyhľadávanie (na základe reprezentovanej otázky 3 sa pomocou odpovede vygenerovanej voči indexu vytvorí zoznam relevantných dokumentov 4) a Usporiadanie (vstupujúci zoznam dokumentov zo 4 je usporiadaný podľa miery relevancie k otázke, výstupom je usporiadaný zoznam 5), kde je potrebné špecifikovať rozdiel medzi štandardným plnotextovým prístupom a

navrhovanou kombináciou s modelom na báze FCA. Usporiadaný zoznam výsledkov je potom odpoveďou na pôvodnú otázku používateľa. Tento proces môže byť doplnený o spätnú väzbu (6) používateľa, čo je opätovného spustenie procesu vyhľadávania s úpravou pôvodnej otázky („query expansion“). Spätná väzba môže byť automatická alebo riadená používateľom, tento vzťah je takisto miestom zmeny oproti štandardnej spätnej väzbe, preto je v schéme zvýraznený.

Logický pohľad na množinu textových dokumentov z hľadiska aplikácie v oblasti IR vytvára model pre indexovanie dokumentov. V klasickom vektorovom modeli pre vyhľadávanie sa zvyčajne používa plnotextové (*full-text*) indexovanie. V našom prípade ide o základnú analýzu textu podobnú procesu predspracovania, aký bol použitý pri konceptuálnej analýze daných množín. Takto je možné použiť klasický vektorový model na báze TF-IDF schémy vo vyhľadávaní, ktorý je zároveň (na základnej úrovni predspracovania) kompatibilný s ďalšou analýzou. Tento model je základnou časťou kombinovaného modelu pre kombinovaný „index“. Jeho druhou časťou je potom práve hybridný model na báze FCA, ktorý bol navrhnutý pre konceptuálnu analýzu množín textových dokumentov. Realizácia „indexu“ je potom založená na vytvorení index súborov plnotextového vyhľadávania a jeho uloženie vo forme súborov na disku (ako serializované objekty), ako aj na analýze danej množiny hybridným algoritmom a uložení jeho výstupu v podobe serializovaného objektu (resp. objektov) zahrňujúc:

- Lokálne FCA modely pre dekomponované podmnožiny dokumentov a ich popis pomocou charakteristických termov s váhami, tzn. jeden uzol pre každý lokálny FCA zväz, spolu so zoznam príslušných dokumentov v uzle (lokálnom zväze) a vektorom dvojíc (charakteristický term t, váha w_t)
- Vybudovaný model hierarchie uzlov spájajúcich FCA modely podľa ich popisov
- Pre prípad využitia ako pomocného prostriedku pre prehľadávanie a vizualizáciu množiny dokumentov, aj štruktúru pôvodného výstupu algoritmu GHSOM použitého pri dekompozícii. Táto zložka je z hľadiska aktuálneho návrhu voliteľná, jej využitie však môže poskytnúť viacero vylepšení a rozšírení systému.

Plnotextový index v kombinácii s touto konceptuálnou štruktúrou (ktorá plní funkciu tzv. ontologickej komponenty) je potom základnou štruktúrou využívanou v procese vyhľadávania v rámci navrhovaného systému. Práve použitie tohto „indexu“ potom určuje rôzne možnosti vyhľadávania výsledkov (podrobný popis je možné nájsť v [3]):

- Plnotextové vyhľadávanie – otázka vo vektorovej reprezentácii sa použije priamo pre vyhľadávanie. V prípade rozšíreného „indexu“ sa použije len zložka odpovedajúca vyhľadávaniu v klasickom indexe na báze vektorového modelu.
- Vyhľadávanie len s použitím konceptuálneho modelu – vektorová reprezentácia otázky sa porovnáva s hybridným FCA modelom. Výsledkom je množina relevantných dokumentov nájdených v rámci FCA modelu vzhľadom k danej otázke.
- Rozšírené plnotextové vyhľadávanie s automatickým expandovaním otázky – pôvodná otázka vo vektorovej reprezentácii sa použije na expandovanie do rozšírenej podoby, táto sa potom použije pri vyhľadávaní v klasickom indexe, pričom expandovanie je založené na vyhľadávaní v konceptuálnom modeli (ako v predchádzajúcom prípade). V tomto prípade ide o kombináciou plnotextového a konceptuálneho vyhľadávania (expandovanie otázky = automatická spätná väzba).
- Kombinované plnotextové a konceptuálne vyhľadávanie – výsledkom vyhľadávania sú dva zoznamy, jeden pre plnotextové vyhľadávanie (s alebo bez expandovania otázky) a jeden pre konceptuálne vyhľadávanie, výsledky sú usporiadané neskôr.

Konceptuálne vyhľadávanie kombinuje vyhľadávanie vhodného konceptového zväzu prehľadávaním aglomeratívnej hierarchickej štruktúry (a to buď striktné alebo čiastočné hierarchické vyhľadávanie) a hľadanie konceptu (či konceptov) v nájdených konceptových zväzoch pre nájdenie dokumentov. Posledným základným prvkom systému je blok pre usporiadanie nájdených dokumentov. V prípade plnotextového prístupu je vyhodnocovacia funkcia priamo postavená na vektorovom modeli IR (podobnosť dokumentov k otázke). Pri konceptuálnom vyhľadávaní usporiadanie získavame na základe zoznamu konceptov s ich extentmi (čo sú zoznamy dokumentov). Pre kombinované vyhľadávanie sú výsledkom dva zoznamy, tie sa môžu ponúknuť samostatne (potom sú usporiadané podľa jednotlivých prístupov) alebo ako kombinovaný zoznam.

IV. NÁVRH METODOLÓGIE PRE ZNOVUPOUŽITIE NAVRHNUTÝCH POSTUPOV

Reálne využitie navrhnutých metód a systémov v širšom alebo novom kontexte (doméne) si vyžaduje určitú metodológiu pre nasadenie, čo (v závislosti od cieľov aplikácie) znamená

napríklad vytvorenie nového konceptuálneho modelu alebo doplnenie existujúceho, zmenu bázy dokumentov alebo úpravu metód používania modelu. Metodologický pohľad by mal byť súčasťou návrhu riešenia pre realizáciu v prostredí sémantického webu. Zavŕšením procesu je teda metodológia popisujúca aplikáciu teoretických a aplikačných prvkov v novej doméne alebo probléme, k tomu však je potrebné pridať niekoľko poznámok:

- Použitý „problém“ (alebo problémová oblasť) je IR s podporou konceptuálneho (ontologického) modelu. V tejto kapitole berieme do úvahy túto aplikáciu, avšak metodológia môže byť použitá na rôzne aplikačné oblasti.
- Najvhodnejšou metodológiou pre tento popis je METHONTOLOGY [4], vzhľadom k tomu, že:
    - má veľký vzťah k znalostnému inžinierstvu a detailný popis
    - je aplikačne nezávislá (z hľadiska stratégie tvorby ontológie)
    - existuje u nej vývoj ontológie ako súboru prototypov (možnosť postupne dopĺňať vytvorený model)
    - neobmedzuje použitú formalizáciu znalostí
    - bola úspešne aplikovaná vo viacerých doménach
    - bola dokonca použitá aj v prípade metód na báze FCA – v [1] autori deklarujú použitie upravenej metodológie kombinujúcej METHONTOLOGY a metódy na extrakciu ontológie pomocou RCA (rozšírený prístup na báze FCA).

Vybranou metodológiou pre súhrn a opätovné použitie navrhnutých metód je teda METHONTOLOGY, ktorú sme doplnili a upravili podľa potreby. Cieľom metodológie je nájsť vyčerpávajúci popis každého konceptu a relácie ontológie. Dôležité je ešte to, že v našom prípade sa zameriavame na automatickú tvorbu modelu, kde nekomunikujú tvorcovia (ľudia), ale model je vytváraný automaticky v zmysle analýzy dát. Preto niektoré kroky pôvodnej metodológie sú vyzdvihnuté, kým iné nie sú také významné (a nemusia byť použité).

Proces tvorby ontológie (alebo teda konceptuálneho modelu na báze FCA) pomocou novej metodológie sa teda odkazuje na aktivity vykonávané počas budovania nasledovne:

- Aktivity projektového manažmentu – manažment nie je definovaný v zmysle kooperácie viacerých ľudí (aspoň nie v zmysle nejakého reálneho pracovného toku), proces reálne riadi administrátor.
    - Plánovanie – identifikuje, ktoré úlohy budú uskutočňované, ako budú usporiadané, koľko času a prostriedkov potrebujeme na ich kompletizáciu, potom v našom prípade máme úlohy:
        - Identifikovanie zdrojov dokumentov a ich predbežná analýza – výber množiny dokumentov, pre ktorú sa vytvára konceptuálny model, určuje sa počet dokumentov, predbežné štatistiky po predspracovaní o počte a rozložení termov, predbežný návrh pre nastavenia parametrov algoritmov GHSOM (tau1, tau2, počet objektov v neexpandovaných podmapách, atď.), FCA (prah minimálnej frekvencie), použité miery podobnosti a schémy pre váhovanie termov, atď.
        - Odhad množstva času a výpočtových prostriedkov pre analýzu množín a tvorbu modelu – na základe skúseností alebo čiastočných odhadov administrátor určí približne čas na spracovanie množiny v rámci celého procesu pri rôznych možnostiach výpočtovej sily. Reálne je potrebné odhadnúť beh algoritmu zhlukovania GHSOM, rozdelenia množín, vybudovania lokálnych FCA modelov a vytvorenie hybridného modelu pomocou aglomeratívneho zhlukovania.
        - Aplikačné odhady – plánovanie aktivít a zdrojov pre potreby problému, v našom prípade systému pre IR (veľkosť plnotextového a konceptuálneho indexu, použité prístupy vyhľadávania, definícia prostriedkov pre uchovávanie indexu a súvisiace vlastnosti).
    - Kontrola – garantuje, že plánované úlohy sú dokončené takým spôsobom, ako boli navrhnuté – v našom prípade máme automatickú tvorbu, kontrola teda môže byť prevedená len ako test uskutočnenia všetkých krokov a splnenia vybraných parametrických rozsahov pre algoritmy.

  - Záruka kvality – zaručuje, že kvalita každého z výstupných produktov (ontológia, softvér, dokumentácia) je dostatočná – v našom prípade úzko prepojené s kontrolou, cieľom je testovanie splnenia navrhnutých odhadov a obmedzení, napríklad pre veľkosť lokálnych množín pri dekompozícii a pod.
- Vývojovo orientované aktivity – aktivity súvisiace priamo s vývojom modelu
  - Špecifikácia – určuje prečo bola ontológia postavená, aké je jej očakávané využitie a kto sú koncoví používatelia – v našom prípade:
    - Cieľ – vytvorenie konceptuálneho modelu na báze FCA popisujúceho vybranú množinu dokumentov
    - Očakávané využitie – použitie vytvoreného modelu v systéme pre IR, ktorý kombinuje konceptuálnu a plnotextovú zložku
    - Koncoví používatelia – dva prípady: 1. z hľadiska programovania – aplikačný programátor systému IR; 2. z hľadiska konečného vyhľadávania – klasický používateľ vyhľadávajúci dokumenty.
  - Konceptualizácia – štruktúra doménových znalostí reprezentovaná ako zmysluplný model na znalostnej úrovni, v klasickej metodológií sa tu extrahuje zo zdrojov slovník termov – v našom prípade, za predpokladu, že doménové znalosti sú reprezentované danou množinou textových dokumentov, je konceptualizáciou v podstate takisto slovník termov, avšak rozšírený o ďalšie veci, to znamená nasledovný súbor objektov (pričom podpornou aktivitou je proces získavania znalostí):
    - Vytvorené lokálne konceptové zväzy – súbor konceptov v rámci zväzov spolu s reláciami na základe usporiadania
    - Extrahované termy popisujúce lokálne zväzy – charakteristické termy popisujúce zväzy s váhami
    - Aglomeratívna hierarchická štruktúra spájajúca popisy zväzov – popisy uzlov hierarchickej štruktúry a ich vzájomné vzťahy
    - Výsledky zhlukovania GHSOM pri dekompozícii – voliteľné, môže slúžiť ako vizualizačná pomôcka, napríklad aj pri vyhodnotení a kontrole
  - Formalizácia – transformuje konceptuálny model do formálneho alebo semi-formálneho modelu – v našom prípade je hybridný model priamo formalizáciou, pretože sa používa ako reprezentácia znalostí v systéme IR (druhá možnosť, ktorú ale neuplatňujeme, je formalizácia ako preklad modelu FCA do niektorého štandardného jazyka ontológie):
    - Formalizácia je spojená s konceptualizáciou, pretože priamo vytvorený hybridný model je formálnou reprezentáciou. Keďže v teórii UI je reprezentácia znalostí definovaná ako súbor údajových štruktúr a operátorov pre prácu s nimi, náš hybridný model a metódy pre prácu s ním v rámci IR sú priamo použitou reprezentáciou znalostí. Ak by sme v rámci metodológie chceli uplatniť zmenu operátorov pre prácu s hybridným modelom FCA, potom formalizácia by bola samostatným krokom, ale len pre túto úpravu (nie FCA modelu).
  - Implementácia – buduje vypočítateľný model v počítačovom jazyku – v našom prípade máme:
    - Výber technológií – v prípade potreby je možné podľa implementačných skúseností (napríklad uvedených aj v tejto práci) realizovať výber vhodných komponentov pre implementáciu.
    - Realizácia systému IR s danou množinou dokumentov
  - Údržba – aktualizuje a opravuje ontológiu – v prípade použitia inkrementálnych algoritmov by bolo možné upravovať model priebežne a dopĺňať nové domény – v našom prípade sa však uvažuje aplikácia v rámci vopred definovanej množiny, takže tento krok sa neuplatňuje. Použitie v rámci viacerých množín naraz je možné s hľadiska IR realizovať ako otázku do viacerých modulov (každá množina samostatne) a ich dodatočnou kombináciou.
- Podporné aktivity
  - Získavanie znalostí – v našom prípade je to základná podporná aktivita,

ktorá je v podstate priamo spojená s procesom konceptualizácie, ako bol pred chvíľou popísaný. Takisto sem patria všetky analýzy a štatistiky súvisiace s predspracovaním množiny dokumentov a získavania odhadov.

- Vyhodnotenie – technické zhodnotenie ontológií, asociovaného softvéru, prostredia a dokumentácie, z dôrazom na každú fázu tzv. životného cyklu – v našom prípade sa zameriava na vyhodnotenie hybridného modelu z pohľadu záruk kvality a kontroly.
- Integrácia ontológií – požaduje sa v prípade ak budovanie ontológie bolo dosiahnuté znovupoužitím existujúcich ontológií – v našom prípade sa neuplatňuje, konceptuálny model vzniká automaticky z dát bez podpory existujúcich ontológií (konceptuálnych modelov).
- Dokumentácia – detailný popis pre každú fázu a vygenerovaný výstup – cieľom by mal byť v našom prípade popis (napríklad vo forme XML dokumentu + serializovaných súčastí modelov):
  - Použitých nástrojov a metód predspracovania
  - Prvotnej analýzy, štatistík a odhadov
  - Vygenerovaných verzií výsledného hybridného modelu na báze FCA (prípadne i submodelov) – spolu s popisom nastavení algoritmov a jeho vyhodnotením z hľadiska kontroly a záruk kvality.
  - Použitých metód pre prácu s modelom v rámci systému IR – nastavenie módov a operátorov vyhľadávania, konkrétne príklady otázok a odpovedí, popis vlastností „indexu“.
- Konfiguračný manažment – zaznamenáva každú verziu dokumentácie, softvéru a ontológie pre kontrolu zmien – v našom prípade to je možnosť použiť verzovanie navrhnutých modelov, bolo by však potrebné uchovávať dokumentačné popisy v nejakom systéme pre verzovanie – ide skôr o voliteľný krok.

- Životný cyklus ontológie – identifikuje množinu stavov, ktorými ontológia prešla počas svojho vývinu – v našom prípade sa životný cyklus neuplatňuje (ako už bolo uvedené pri integrácii, zmysel by mal pri inkrementálnej tvorbe a dopĺňaní modelu), model sa vytvára raz a použitie vyhľadávania vo viacerých množín naraz je metodologicky nezávislé.

Popísané kroky a detaily spolu vytvárajú všeobecný pohľad na aplikáciu v rámci ľubovoľnej množiny dokumentov, je to v podstate úvodný skelet. V budúcnosti by bolo potrebné realizovať konkrétny systém IR na báze navrhnutých metód a doplniť metodológiu o praktické údaje a dokumentáciu, ako sú presné definície operátorov vyhľadávania, praktické skúsenosti s nastavovaním parametrov rôznych algoritmov a vyhodnotenie vytvorených modelov, ako aj testy kvality vyhľadávania v rámci danej množiny (domény). V prípade voľby aplikácie v inom probléme (iné ako IR) by bolo potrebné riešiť aj všeobecnú náplň krokov, pretože by pravdepodobne bolo nutné zmeniť napríklad spôsob manipulácie s modelom, či dokonca jeho štruktúru a obsah.

POĎAKOVANIE

Táto práca bola podporovaná KEGA grantom MŠVVaŠ SR č. 065TUKE-4/2011.

LITERATÚRA

[1] Bendaoud et al, “Text-based ontology construction using relational concept analysis,” in Proceedings of the International Workshop on Ontology Dynamics, Innsbruck, Austria, 2007, pp. 55-68.

[2] P. Butka, “Jednoduchá metóda kombinácie lokálnych modelov pri budovaní hierarchie konceptov z textových dokumentov,” in Znalosti 2007, Proceedings of the 6th annual conference, VŠB-TU Ostrava, Czech Republic, February, 2007, pp.167-178.

[3] P. Butka, J. Pócsová, “Hierarchical FCA-based conceptual model of text documents used in information retrieval system,” in SACI 2011, 6th IEEE International Symposium on Applied Computational Intelligence and Informatics, IEEE conference, 2011, pp. 199-204.

[4] B. Ganter, R. Wille, *Formal Concept Analysis: Mathematical Foundations.* Springer, 1998.

[5] A. Gomez-Perez, M. Fernandez-Lopez, A. de Vicente, “Towards a Method to Conceptualize Domain Ontologies,” in Proceedings of ECAI96 Workshop on Ontological Engineering, Budapest, Hungary, 1996, pp. 41-51.

Porovnanie metodológií pre budovanie ontologických modelov

Peter BUTKA

Centrum pre informačné technológie, Katedra kybernetiky a umelej inteligencie, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská Republika

peter.butka@tuke.sk

***Abstrakt* — Táto práca sa zaoberá popisom vybraných metodológií pre budovanie ontológií v praxi a ich porovnaním. Odlišnosť použitia jednotlivých metodológií je prevažne v praktických aspektoch konceptuálneho modelovania, ako aj v doméne použitia. V závere práce je v krátkosti zhrnutých niekoľko odporúčaní z hľadiska vhodnosti použitia pre potreby budovania konceptuálneho modelu na základe konceptov a vzťahov získaných procesmi na báze metód dolovania znalostí z textových dokumentov.**

***Kľúčové slová* — ontológie, metodológie tvorby ontológií, konceptuálne modely, porovnanie**

I. Úvod

Viacero výskumných skupín navrhlo postupnosť krokov (metodológiu) pri budovaní ontológií, každá však zväčša navrhla svoj vlastný špecifický prístup. Na základe skúseností získaných pri vývoji Enterprise Ontology [7] a projektu TOVE (TOronto Virtual Enterprise) [4] (obidva v doméne podnikového modelovania) bola podrobne popísaná a prezentovaná prvá z takýchto metodológií, neskôr prepracovaná v [6] a [8]. Následne Bernaras v [1] prezentoval metodiku budovania ontológie v doméne elektrických sietí ako súčasť Esprit KACTUS projektu. V [3] autori predstavili METHONTOLOGY. Poslednou spomenutou bude metodika použitá pri budovaní SENSUS ontológie [5].

V rámci tejto práce sa zameriame najprv na popis jednotlivých metodológií, neskôr budú tieto porovnané podľa rôznych kritérií, a na záver uvedieme niektoré odporúčania z ohľadom na možnosti použitia vhodnej metodológie pre potreby budovania konceptuálneho modelu a jeho praktickú aplikáciu, najmä z pohľadu sémantických aplikácií (v rámci napĺňania myšlienky sémantického webu) v doméne vyhľadávania informácií z podporou dolovania znalostí z textov.

II. Popis vybraných metodológií pre budovanie ontológií

V tejto kapitole si popíšeme vybrané metodológie spomenuté v úvode, jednotlivé referencie na príslušné metodológie boli takisto uvedené v úvode.

Uschold & King

Táto metodológia je postavená na skúsenostiach pri budovaní Enterprise Ontology, ontológie pre modelovanie procesov v podnikoch („*enterprise modelling*“). Samotná metodológia je popísaná nasledujúcimi inštrukciami:

1. *Identifikácia účelu* – definícia prečo sa ontológia vytvára a aké je jej budúce použitie.
2. *Budovanie ontológie* (rozdelené na tri kroky):
    i. *Zachytenie ontológie*, tzn.:
        - Identifikácia kľúčových konceptov a ich vzťahov v danej doméne. Je dôležité sústrediť sa na podstatu konceptov tak ako sú, nie na slová, ktoré ich reprezentujú.
        - Tvorba presných jednoznačných textových definícií pre takéto koncepty a vzťahy.
        - Identifikácia termov, ktoré majú dané koncepty a vzťahy referovať.

    Autori pre dosiahnutie tohto kroku používajú tzv. middle-out

prístup.

ii. *Kódovanie* – jej výstupom je explicitná reprezentácia znalostí získaných v predchádzajúcom kroku vo formálnom jazyku.

iii. *Integrácia existujúcich ontológií* – v prípade oboch predchádzajúcich krokov je to miesto pre zváženie použitia existujúcich ontológií na dosiahnutie špecifických cieľov.

3. *Vyhodnotenie* – zhodnotenie technických detailov ontológie, asociovaného softvéru a dokumentácie s dôrazom na požadované špecifikácie a reálne aspekty používania.
4. *Dokumentácia*.

Najvýznamnejším projektom vytvoreným pomocou tejto metodológie je Enterprise Ontology (množina termov a definícií relevantných k podnikom a obchodným firmám). V rámci toho bol vytvorený Enterprise Toolset, agentová architektúra zahŕňajúca Procedure Builder (pre zachytenie procesných modelov), Agent Toolkit (podpora vývoja agentov), Task Manager (integrácia, vizualizácia a podpora pre získavanie a spracovanie procesov a úloh) a samotná ontológia *Enterprise Ontology* ako komunikačný prostriedok.

Gruninger & Fox

Táto metodológia je založená na skúsenostiach z projektu tvorby ontológie TOVE, opäť na tej istej doméne modelovania procesov a aktivít v podnikoch. V podstate si vyžaduje vybudovanie logického modelu znalostí, ktoré sú špecifikované prostredníctvom ontológie. Tento model nie je konštruovaný priamo, najprv je špecifikovaný neformálny popis a potom je formalizovaný. Jednotlivé kroky je možné nasledovne popísať:

1. *Zachytenie motivačných scenárov* – podľa autorov je tvorba ontológií motivovaná scenármi, ktoré sa objavujú z možnej aplikácie. Motivačné scenáre sú problémy alebo príklady, ktoré nie sú adekvátne pokryté existujúcimi ontológiami, každý scenár predkladá množinu možných riešení týchto problémov (riešenia sú poskytnuté zatiaľ pomocou neformálnej sémantiky pre objekty a relácie, ktoré sú neskôr vkladané do ontológie). Každý návrh novej ontológie, alebo rozšírenia existujúcej musí byť popísaný jedným alebo viacerými motivačnými scenármi, ako aj množinou ich možných riešení.
2. *Formulácia neformálnych kompetenčných otázok* – tieto sú založené na scenároch získaných v predchádzajúcom kroku, sú to špecifické otázky v prirodzenom jazyku hľadajúce požiadavky vo forme otázok. Ontológia musí byť schopná reprezentovať tieto otázky použijúc zvolenú terminológiu, pričom je potreba schopnosti charakterizovať tieto otázky pomocou axióm a definícií vo formálnom jazyku. Otázky slúžia ako ohraničenia toho, čím ontológia môže byť. Neexistuje jediná ontológia asociovaná s množinou otázok, tieto sú použité na vyhodnotenie ontologických záväzkov v momente, kedy ontológia spĺňa množinu požiadaviek.
3. *Špecifikácia terminológie ontológie vo formálnom jazyku*:
    a. *Spracovanie neformálnej terminológie* – z množiny otázok je extrahovaná množina termov, ktoré slúžia ako základ pre špecifikáciu terminológie vo formálnom jazyku.
    b. *Špecifikácia formálnej terminológie* – termy získané na základe predchádzajúceho kroku sa transformujú do formalizmu, tieto termy vytvoria definície a ohraničenia neskôr vyjadrené pomocou axióm.
4. *Formulácia formálnych otázok použitím terminológie ontológie* – použitím formálnych termov sa transformujú otázky.
5. *Špecifikácia axióm a definícií použitím termov ontológie vo formálnom jazyku* – axiómy v ontológii špecifikujú definície v termoch ontológie a ohraničenia na základe ich interpretácie (sú definované ako logické tvrdenia prvého rádu). Avšak jednoduché vymenovanie termov v logike prvého rádu nestačí, nevytvára to ontológiu, axiómy musia definovať sémantiku termov. Proces tvorby axióm je iteračný, v každom kroku pridávame potrebné prvky preto, aby sme pokryli kompetenčné otázky a návrhy riešení motivačných scenárov.
6. *Ustanovenie podmienok pre charakterizáciu úplnosti ontológie* – potom čo boli kompetenčné otázky formálne zapísané, musíme definovať podmienky pre dosiahnutie úplného riešenia.

Táto metodológia bola použitá pri tvorbe ontológií v rámci TOVE (Toronto Virtual Enterprise), ako sú Enterprise Design Ontology, Project Ontology, Scheduling Ontology a Service Ontology. Na základe tejto metodológie boli vybudované ontológie pre rôzne aplikácie, napr.:

- prostredie pre návrh procesov a aktivít v rámci oblasti modelovania podnikových procesov, ako navrhovanie, analýza, či re-dizajn, kde aplikácia ontológie spočíva v komparatívnej analýze navrhovaných alternatív a v asistentskej podpore návrhu.
- integrovaná agentová podpora pre manažment dodávateľských reťazcov, v rámci ktorej sa organizovala sieť inteligentných kooperujúcich agentov (jednotlivé prvky dodávateľských reťazcov, plnia jednu alebo viac dodávateľských funkcií). TOVE virtuálny podnik poskytol testovacie prostredie pre otestovanie agentov pri plnení logistických, transportných a manažérskych úloh.

Bernaras (KACTUS)

Jedným zo základných cieľov Esprit KACTUS projektu je preskúmanie znovupoužiteľnosti znalostí v komplexných technických systémoch a funkcie ontológií pre ich podporu. Tento prístup je podmienený vývojom aplikácie (vždy pri budovaní aplikácie ontológia reprezentuje znalosti vyžadované pre jej tvorbu). Ontológia tak môže byť vytvorená znovupoužitím a integráciou ontológií zo skorších aplikácií, jednotlivé kroky sú:

1. *Špecifikácia aplikácie* – aplikačný kontext a prehľad komponentov budúcej aplikácie.
2. *Predbežný návrh založený na relevantných top-level ontologických kategórií* – zoznamy termov a úloh vytvorených v predchádzajúcej fáze sa použijú ako vstupy pre získanie pohľadov na globálny model vo vzťahu k determinovaným top-level ontologickým kategóriám. Tento proces uprednostňuje vývoj ontológií použitím iných existujúcich , ktoré sú upravené a rozšírené pre použitie v novej aplikácií.
3. *Úprava a štruktúrovanie ontológie* – môže byť použitý princíp minimálneho spájania na dosiahnutie čo najväčšej nezávislosti modulov navzájom, čím získame koherentnejšiu ontológiu (maximalizujeme homogenitu modulov).

Ako skúsenosti získané použitím tohto postupu, autori prezentujú vývoj troch ontológií ako výsledok vývoja troch aplikácií (diagnostika chýb v elektrickej sieti, plánovanie obnovy služieb po chybe v elektrickej sieti, kontrola elektrickej siete na báze predchádzajúcich dvoch aplikácií).

METHONTOLOGY

Táto metodológia umožňuje konštrukciu ontológií na znalostnej úrovni zahrňujúc identifikáciu procesu pre vývoj ontológií, životný cyklus založený na vývine prototypov a konkrétne techniky pre vykonanie každej aktivity.

Proces tvorby ontológie sa tu odkazuje na aktivity vykonávané počas budovania ontológie. Kritické je identifikovať tieto aktivity za účelom potvrdenia zhody pri ontológiách budovaných geograficky oddelenými, dištančne kooperujúcimi tímami s určitou zárukou korektnosti a úplnosti. Autori uvádzajú tri kategórie takýchto aktivít:

- *aktivity projektového manažmentu*
  - *plánovanie* – identifikuje, ktoré úlohy budú uskutočňované, ako budú usporiadané, koľko času a prostriedkov potrebujeme na ich kompletizáciu.
  - *kontrola* – garantuje, že plánované úlohy sú dokončené takým spôsobom, ako boli navrhnuté.
  - *záruka kvality* – zaručuje, že kvalita každého z výstupných produktov (ontológia, softvér, dokumentácia) je dostatočná.
- *vývojovo orientované aktivity*
  - *špecifikácia* – určuje prečo bola ontológia postavená, aké je jej očakávané využitie a kto sú koncoví užívatelia.
  - *konceptualizácia* – štruktúra doménových znalostí reprezentovaná ako zmysluplný model na znalostnej úrovni.
  - *formalizácia* – transformuje konceptuálny model do formálneho alebo semi-formálneho modelu.
  - *implementácia* – buduje vypočítateľný model v počítačovom jazyku.
  - *údržba* – aktualizuje a opravuje ontológiu.
- *podporné aktivity*
  - *získavanie znalostí* – extrahuje znalosti z danej domény.
  - *vyhodnotenie* – technické zhodnotenie ontológií, asociovaného softvéru, prostredia a dokumentácie, z dôrazom na každú fázu tzv. *životného cyklu*.
  - *integrácia ontológií* – požaduje sa v prípade ak budovanie ontológie bolo dosiahnuté znovupoužitím existujúcich ontológií.
  - *dokumentácia* – detailný popis pre každú fázu a vygenerovaný výstup.
  - *konfiguračný manažment* – zaznamenáva každú verziu dokumentácie,

softvéru a ontológie pre kontrolu zmien.

Životný cyklus ontológie identifikuje množinu stavov, ktorými ontológia prešla počas svojho vývinu, opisuje aké aktivity boli použité v každom stave a aké vzťahy medzi jednotlivými stavmi sú (prioritné vzťahy, návraty, atď.). Životný cyklus je založený na iterácii a evolúcii prototypov.

Z ontológií, ktoré vznikli pomocou METHONTOLOGY, spomeňme ontológiu o doméne chemických prvkov a kryštalických štruktúr, o látkach znečisťujúcich environmentálne prostredie (metódy detekcie rôznych látok v rôznych prostrediach), prípadne ontológiu referencií na iné ontológie (niečo ako „žlté stránky" existujúcich ontológií) a reštrukturovaná verzia ontológie zhromažďujúcej informácie o výskume v oblasti získavania znalostí („*knowledge acquisition*"). Informácie obsiahnuté v týchto ontológiách potom využívajú rôzne aplikácie.

SENSUS

SENSUS je ontológia pre použitie v spracovaní prirodzeného jazyka, predstavujúca konceptuálnu štruktúru pre vývoj strojových prekladačov. Jej aktuálny obsah bol získaný extrakciou a spájaním informácií z rôznych elektronických zdrojov znalostí. Na začiatku boli prepojené dve vysoko-úrovňové lingvistické ontológie (PENMAN Upper Model a ONTOS) a ručne sémantické kategórie zo slovníka na vytvorenie ontologickej bázy. Potom bol (opäť ručne) s bázou spojený WordNet a výsledok bol pomocou špecifického nástroja spojený s anglickým slovníkom. Nakoniec, pre podporu strojového prekladu, boli pripojené španielske a japonské lexikálne vstupy z prekladových slovníkov. Vďaka celému procesu má SENSUS viac ako 50000 konceptov organizovaných v hierarchii (sú z množiny vysokej a strednej úrovne abstrakcie, špecifické domény nie sú pokryté). Pre tvorbu konkrétnej ontológie pre špecifickú doménu môžeme postupovať nasledovne:

1. Skupina termov je vybraná ako počiatočná („*seed*").
2. Tieto termy sú ručne nalinkované na SENSUS.
3. Všetky koncepty na ceste od počiatočného termu ku koreňu sú vložené.
4. Termy, ktoré sú relevantné k danej doméne (a ešte sa neobjavili) sú pridané.
5. Nakoniec je pre uzly, cez ktoré vedie veľa ciest, niekedy pridaný celý podstrom daného uzla (toto sa deje manuálne, pretože rozhodnutie si vyžaduje rozumieť doméne). Opatrnosť je na mieste, veľa vysoko-úrovňových uzlov v ontológií je vždy na množstve ciest, ale je ťažko možné, že celý podstrom bude patriť do danej domény.

Pomocou tejto metodológie bola vytvorená ontológia pre plánovanie vojenských vzdušných akcií (boli charakterizované základné elementy ako plán akcie, akcia, scenár, účastník, veliteľ, atď.). Táto ontológia taktiež zahŕňa znalosti o zbraniach, palubných systémoch, palive a pod. Príklad aplikácie na báze tejto ontológie je napr. *Strategy Development Assistant*, nástroj pre podporu inteligentného a asistovaného návrhu plánov.

III. POROVNANIE VYBRANÝCH METODOLÓGIÍ

Autori nových ontológií zvyčajne dodržiavajú postupy získané analýzou práve tejto skupiny metodológií, preto považujem ich množinu za dostatočnú vzorku pre porovnanie a analýzu. Práve úprava jednej z nich by mala viesť k návrhu očakávanej metodológie pre tvorbu a využitie konceptuálnych modelov v sémantickom webe v rámci vyhľadávania informácií.

Existuje množstvo rôznych pohľadov na metodológie a spôsobov hodnotenia krokov, z ktorých sa skladajú, napríklad podľa [2] to sú:

- *vzťah k oblasti znalostného inžinierstva* – zváženie vplyvu „klasického" znalostného inžinierstva na metodológiu.
- *detailnosť metodológie* – kritérium hodnotiace ako exaktne sú aktivity a techniky použité danou metodológiou špecifikované.
- *formalizácia znalostí* – použité resp. navrhnuté formalizmy pre reprezentáciu znalostí.
- *stratégia budovania ontológie* – ktorá z nasledujúcich stratégií bola použitá:
  - *aplikačne závislá* – ontológia je budovaná ako základ aplikácie nad znalostnou bázou.
  - *aplikačne stredne závislá* – možné scenáre použitia ontológie sú identifikované v kroku špecifikácie.
  - *aplikačne nezávislá* – proces je úplne nezávislý od určeného miesta v znalostnej báze, kam bude v budúcnosti ontológia nasadená.
- *stratégia identifikácie konceptov* – existujúce stratégie je možné rozdeliť

nasledovne [8]:

- *zdola-nahor* (bottom-up) – od najšpecifickejších konceptov k najvšeobecnejším konceptom.
- *zhora-nadol* (top-down) – od najvšeobecnejších konceptov k najšpecifickejším konceptom.
- *„od stredu“* (middle-out) – od najrelevantnejších konceptov (pohybujúcich sa niekde uprostred vzhľadom na úroveň abstrakcie) oboma smermi, t.j. k najvšeobecnejším aj najšpecifickejším konceptom.

- *životný cyklus* – analýza toho, či metodológia implicitne alebo explicitne predkladá odporúčania pre zavedenie určitého životného cyklu.
- *odporúčané techniky a postupy* – špecifikácia toho, či boli navrhnuté zvláštne techniky pre vykonávanie rôznych aktivít a procesov, z ktorých je metodológia zložená.
- *aplikácie metodológie* – ktoré ontológie boli vyvinuté pomocou danej metodológie a aké systémy boli na základe týchto ontológií postavené.

Na základe kritérií pre porovnanie metodológií je možné vlastnosti jednotlivých metodológií zhrnúť do nasledujúcej tabuľky:

Tabuľka 1
Porovnanie metodológií budovania ontológií podľa rôznych kritérií

Metodológia	*Vzťah k znalost. inžinier.*	*Detailnosť metodológie*	*Odporúčaná formalizácia znalostí*	*Stratégia budovania ontológie*	*Stratégia identifikácie konceptov*	*Životný cyklus*	*Odpor. techniky*	*Aplikácie metodológie*
Uschold, King	čiastočný	veľmi nízka	nie	aplikačne nezávislá	middle-out	nie	Nie	iba jedna doména
Gruninger, Fox	malý	nízka	logika	aplikačne stredne závislá	middle-out	detailne popísaný	Nie	iba jedna doména
Bernaras	veľký	veľmi nízka	nie	aplikačne závislá	zhora-nadol	nie	Nie	iba jedna doména
METH-ONTOLOGY	veľký	vysoká	nie	aplikačne nezávislá	middle-out	vývin prototy-pov	takmer vš. akt.	viacero domén
SENSUS	žiadny	stredná	sémantické siete	aplikačne stredne závislá	nešpecifik.	detailne popísaný	Nie	viacero domén

IV. ZÁVEREČNÉ POZNÁMKY A MOŽNOSTI POUŽITIA VYBRANEJ METODOLÓGIE PRE AUTOMATICKY GENEROVANÉ ONTOLÓGIE V KONTEXTE SÉMANTICKÉHO WEBU

Na záver porovnania metodológií ešte niekoľko poznámok:

- postupy nie sú jednotné, každá skupina prichádza s vlastnou metodológiou (aj keď k unifikácii prvých dvoch pristúpili Uschold a Gruninger v [8])
- SENSUS je od ostatných prístupov výrazne odlišnou metodológiou
- pri použití určitého skeletu znovupoužitím existujúcich ontológií sa vhodne podporuje interoperabilita existujúcich systémov

Reálne využitie sémantických metód a systémov v širšom alebo novom kontexte (doméne) si vyžaduje určitú metodológiu pre nasadenie, čo (v závislosti od cieľov aplikácie) znamená napríklad vytvorenie nového konceptuálneho modelu alebo doplnenie existujúceho, zmenu bázy dokumentov alebo úpravu metód používania modelu. Metodologický pohľad by mal byť súčasťou návrhu riešenia pre realizáciu v prostredí sémantického webu. Preto vhodný spôsob napĺňania myšlienky sémantického webu najmä z pohľadu aplikácie konceptuálnych modelov v procese vyhľadávania informácií by mal zahŕňať výber aplikáciu metodológie.

Ako najvhodnejšia metodológiou (použiteľná ako skelet) pre popis krokov budovania konceptuálnych modelov na báze modelov vytvorených v rámci dolovania znalostí z textov sa ukazuje METHONTOLOGY, nakoľko má silný vzťah k znalostnému inžinierstvu, detailný popis, je aplikačne nezávislá (z hľadiska stratégie tvorby ontológie), podporuje vývoj prototypov a bola existuje u nej vývoj ontológie ako súboru prototypov (možnosť postupne dopĺňať vytvorený model), neobmedzuje použitú formalizáciu znalostí a bola úspešne aplikovaná vo viacerých doménach (aj s použitím metód dolovania znalostí z textov). Cieľom metodológie je nájsť vyčerpávajúci popis každého konceptu a relácie ontológie. Navyše ide o semi-automatickú metodológiu, z výraznou podporou automatických prostriedkov v zmysle analýzy dát. Logickým cieľom pre ďalšiu prácu je aplikácia METHONTOLOGY (po určitej

úprave) na konkrétny problém v oblasti vyhľadávania informácii z podporou automatickej tvorby sémantického modelu použitého v samotnom procese vyhľadávania.

POĎAKOVANIE

Táto práca bola podporovaná KEGA grantom MŠVVaŠ SR č. 065TUKE-4/2011.

LITERATÚRA

[1] A. Bernaras, I. Laregoiti, J. Corera, "Building and Reusing Ontologies for Electrical Network Applications," in Proceedings of the European Conference on Artificial Intelligence (ECAI'96), 1996, pp. 298-302.
[2] M. Fernandez-Lopez, "Overview Of Methodologies For Building Ontologies," in Proceedings of the IJCAI-99 workshop on Ontologies and Problem-Solving Methods (KRR5), Stockholm, Sweden, 1999.
[3] A. Gomez-Perez, M. Fernandez-Lopez, A. de Vicente, "Towards a Method to Conceptualize Domain Ontologies," in Proceedings of ECAI96 Workshop on Ontological Engineering, Budapest, Hungary, 1996, pp. 41-51.
[4] M. Gruninger, M. Fox, "Methodology for the design and evaluation of ontologies," in Proceedings from Workshop on Basic Ontological Issues in Knowledge Sharing, Montreal, Canada, 1995.
[5] B. Swartout, R. Patil, K. Knight, T. Russ, "Toward Distributed Use of Large-Scale Ontologies," in Proceedings of the AAAI Symposium on Ontological Engineering, Stanford, USA, 1996.
[6] M. Uschold, "Building Ontologies: Towards A Unified Methodology," in Proceedings of 16th Annual Conference of the British Computer Society Specialist Group on Expert Systems, Cambridge, UK, 1996, pp. 16-18.
[7] M. Uschold, M. King, "Towards a Methodology for Building Ontologies," in Proceeding of the Workshop on Basic Ontological Issues in Knowledge Sharing, Edinburgh, UK, 1995.
[8] M. Uschold, M. Gruninger, "Ontologies: principles, methods and applications," *The Knowledge Engineering Review*, vol.11, no.2, 1996, pp. 93-136.
[9] M. Uschold, M. King, S. Moralee, Y. Zorgios, "The Enterprise Ontology," *The Knowledge Engineering Review*, vol.13, 1998, pp. 31-89.

Experimentálna identifikácia nelineárneho dynamického systému pomocou IDENT Tool v prostredí Matlab

Jakub ČERKALA, Anna JADLOVSKÁ

Katedra kybernetiky a umelej inteligencie, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

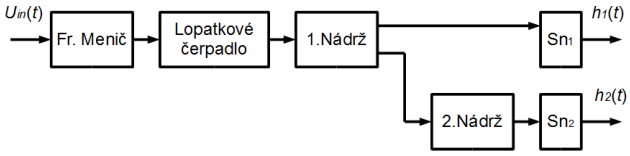
jakub.cerkala@student.tuke.sk, anna.jadlovska@tuke.sk

***Abstrakt* — V tomto článku sa budeme venovať experimentálnej identifikácii s efektívnym nástrojom IDENT Tool, ktorý ponúka System identification toolbox v Matlab-e. Ako vyšetrovaný systém sme zvolili nelineárny simulačný model dvoch otvorených valcových nádrží zapojených za sebou bez interakcie. Simulačným experimentom získame experimentálne dáta, ktoré umelo zašumíme. Následne ich využijeme pre získanie lineárnej aproximácie stochastickým modelom. Celý postup experimentálnej identifikácie vykonáme nástrojom IDENT tool, pričom využijeme viacero aproximačných modelov pre priame porovnanie výsledkov.**

***Kľúčové slová* — experimentálna identifikácia, hydraulický systém, Matlab, stochastické modely**

I. ÚVOD

Koncepcia experimentálnej identifikácie pracuje s dátami a len malým množstvom informácií týkajúcich sa vyšetrovaného objektu. V tomto prípade bude identifikovaný simulačný nelineárny model hydraulického systému, ktorý je získaný analytickou identifikáciou pre parametre podobné reálnemu laboratórnemu modelu, ktorý máme k dispozícii na katedre v Laboratóriu mechatronických systémov V142 [1]. Získané dáta budú umelo zašumené, pričom budeme simulovať realistické prejavy šumu, podobne ako pri laboratórnom modeli.



Obr. 1 Bloková schéma laboratórneho hydraulického modelu

Analytickou identifikáciou získame diferenciálne rovnice nelineárneho modelu. Na základe týchto rovníc naprogramujeme v prostredí Simulink simulačný model, pre ktorý vykonáme experiment budenia vhodným budiacim signálom za účelom získania dátového súboru podobne ako u laboratórneho modelu. Takto získané dáta uložíme do IDDATA objektu, ktorý bude jediným vstupom IDENT tool [2]. Proces predspracovania dát a identifikácia bude vykonaná v tomto grafickom rozhraní.

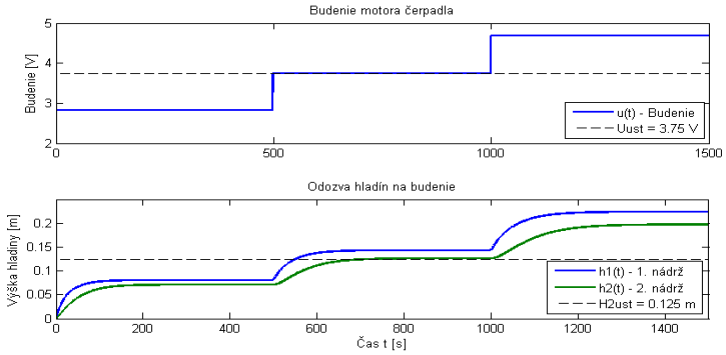
Identifikácia, aproximácia modelom lineárnej štruktúry a následná validácia bude na už základe upravených dát. Výsledné aproximácie modelmi lineárnej štruktúry budú vyčíslené a vzájomne porovnané. Najvhodnejšie aproximácie poslúžia pre syntézu riadenia a taktiež budú graficky porovnané a vyhodnotené.

II. NELINEÁRNY HYDRAULICKÝ MODEL

Parametre nelineárneho modelu budú pre nás ako vyšetrovateľov neznáme. Vieme len odhadnúť rád modelu – predpokladáme model druhého rádu, teda dve nádrže v sérii. Vplyv

čerpadla je pre jednoduchosť zanedbaný.

Predpokladáme, že chceme riadiť výšku hladiny v druhej nádrži $h_2(t)$. Preto musíme vypočítať ustálené budenie čerpadla, ktoré zabezpečuje ustálený prítok do prvej a následne do druhej nádrže. Analyticky by sme to vedeli presne určiť, no vzhľadom na neznáme parametre to nie je možné. Najjednoduchejší spôsob určenia pracovného bodu je postupné zvyšovanie budenia po malých skokoch a sledovanie ustálenia hladiny.

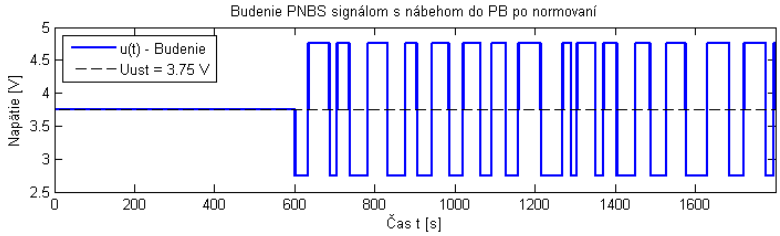


Obr. 2 Experiment určenia pracovného bodu skokovými zmenami budenia

Takto sme určili požadovanú hodnotu H_{2ust} = 0.126 m pre ustálené budenie *Uust* = 3.75 V. Zároveň sme určili limity budenia, pre ktoré nám nevytečú alebo nepretečú nádrže.

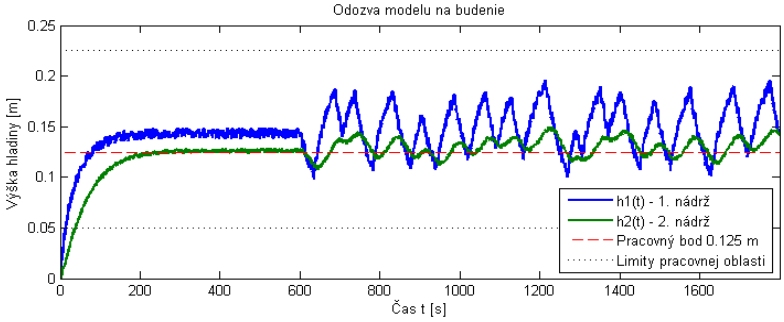
III. Experiment získania dátového súboru pre identifikáciu

Aby sme pokryli čo najväčší rozsah dynamického správania sa identifikovaného modelu, využijeme pseudonáhodný binárny signál. Zároveň predpokladáme, že systém je v nulových počiatočných podmienkach a preto pridávame konštantné budenie, pre prvú tretinu času simulácie.



Obr. 3 Referenčný pseudonáhodný budiaci signál s nábehom do pracovného bodu

Takto získame dáta odozvy, ktorá sa pohybuje v okolí zvoleného pracovného bodu a zároveň nepresahuje limity modelu. K nameraným dátam pridávame náhodný šum. Nameranú odozvu výšky hladiny $h_2(t)$ umelo zašumíme náhodnou hodnotou z rozsahu ± 2 mm.



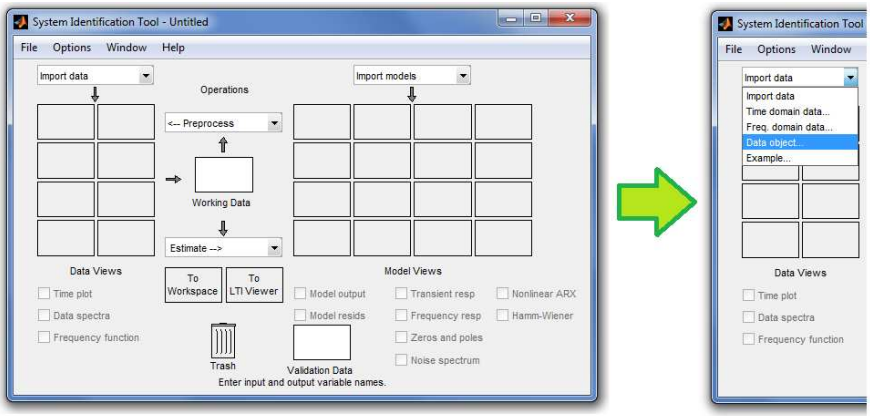
Obr. 4 Odozva hladín nádrži na referenčný signál po umelom zašumení

Takto sme získali dátový súbor, na základe ktorého vykonáme experimentálnu identifikáciu parametrického modelu. Taktiež je pre efektivitu práce vhodné vytvoriť IDDATA objekt experimentálnych dát, kde je možné zadefinovať názvy premenných, periódu vzorkovania či

jednotky skúmaných veličín. Zároveň je takto vytvorený objekt výhodné zobrazovať grafom pomocou príkazu ***idplot***.

IV. IMPORTOVANIE DÁT DO IDENT TOOL

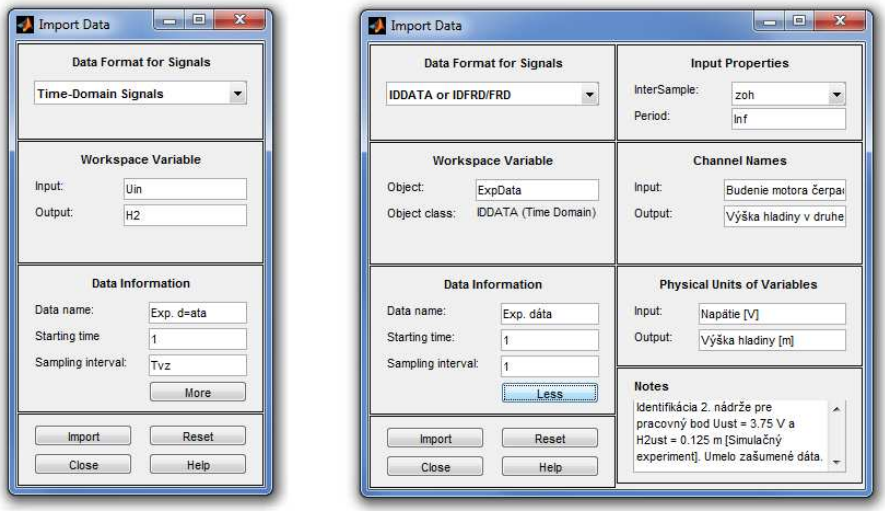
Nástroj IDENT tool spustíme príkazom ***ident*** v príkazovom riadku Matlab-u. Otvorí sa nám grafické rozhranie a môžeme importovať dáta [3].



Obr. 5 IDENT tool – grafické rozhranie pre experimentálnu identifikáciu. Vpravo výber zdroja importu dát

Ak sme si nepripravili IDDATA objekt, môžeme importovať aj vektory dát cez voľbu ***Import data / Time domain data ...*** V tomto prípade však musíme zadať vektor budiaceho signálu $u(t)$ do poľa ***Input***, vektor odozvy hladiny v druhej nádrži $h_2(t)$ ako ***Output***. Okrem toho je dôležité pomenovať tento súbor dát v poli ***Data name*** a nezabudnúť na periódu vzorkovania ***Sampling interval***. Ak by sme chceli doplniť ďalšie informácie, po kliknutí na tlačidlo ***More*** sa nám rozhranie rozšíri.

Ak importujeme IDDATA objekt - ***Import data / Data object ...***, stačí napísať jeho meno a všetko je automaticky vyplnené.



Obr. 6 Vľavo rozhranie pre import dát po vektoroch, vpravo import IDDATA objektu s výpisom doplnkových informácii

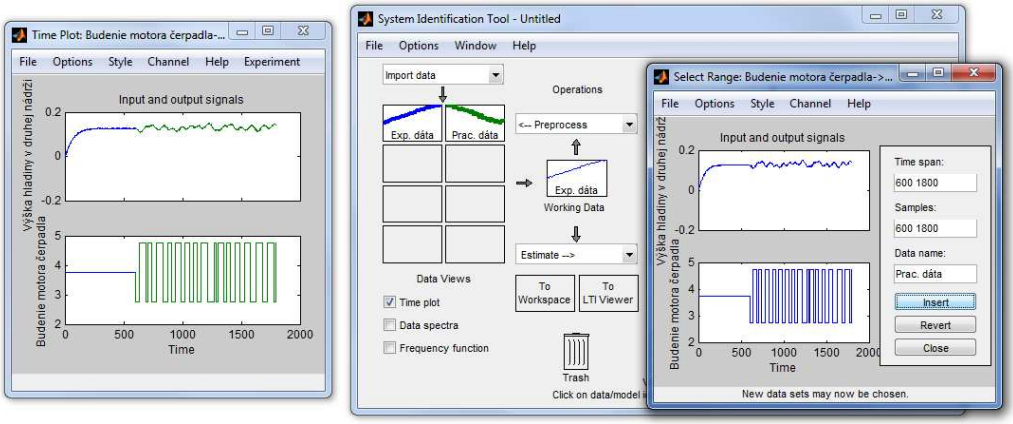
Ak sme postupovali správne, po potvrdení tlačidlom ***Import*** a zatvorením okna v hlavnom okne (Obr. 5) sa nám objaví dátový súbor *Exp. dáta*.

V. PREDSPRACOVANIE DÁT NÁSTROJOM IDENT TOOL

A. Odstránenie nábehu

Prvým krokom predspracovania je odstránenie nábehu. Nábeh sme volili pre prvých 599 vzoriek budiaceho signálu. V hlavnom okne musí byť ako ***Working data*** označený dátový súbor *Exp. dáta*. Potom voľbou ***Preprocess / Select range ...*** otvoríme okno, kde si vyberáme rozsah nových dátových súborov. Chceme iba dáta od vzorky 600 a nazveme ich *Prac. dáta*.

V hlavnom okne máme možnosť zobrazovať dáta v grafoch. V oblasti ***Data Views*** zaškrtneme ***Time plot*** a otvorí sa nám IDPLOT pre vybraté dátové súbory. Výber dátových súborov, ktoré sa majú v grafoch zobrazovať je označený hrubou čiarou v okne setu. Po kliknutí na dátový súbor sa vyberie alebo nie.

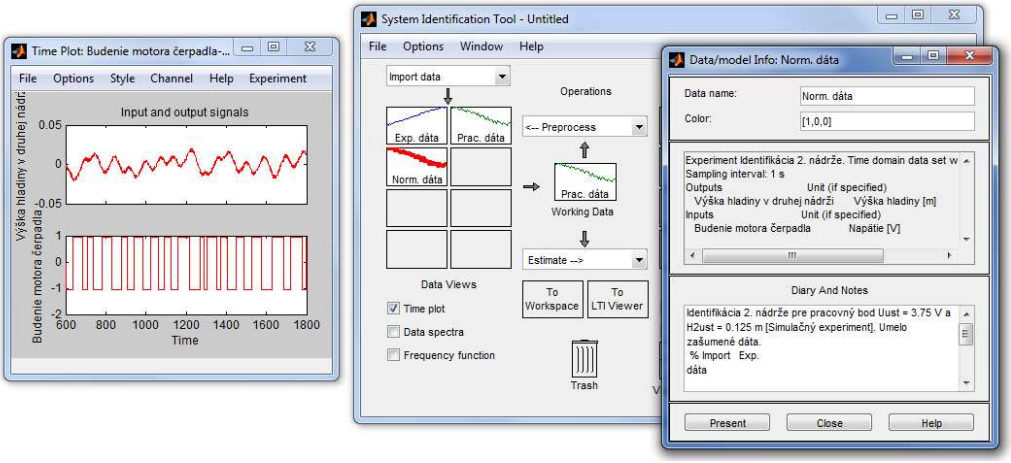


Obr. 7 Vľavo graf experimentálnych a pracovných dát, v strede hlavné okno s dátovými setmi a vpravo rozhranie pre výber z rozsahu dát

B. Normovanie pracovných dát

V ďalšom kroku musíme normovať dáta. Teraz však už pracujeme s dátovým súborom *Prac. Dáta*, preto ho ťahom myši presunieme do okna ***Working Data***. Podobne môžeme poslať dáta do Matlab-u (***To Workspace***) alebo do koša (***Trash***). Normovanie znamená odstránenie konštantných hodnôt. Vzhľadom na to, že sme pre identifikáciu používali pseudonáhodný binárny signál, ktorý sa preklápal okolo ustálenej hodnoty, jeho priemerom teda musí byť daná ustálená hodnota. To isté platí aj pre odozvu.

Tento úkon vykonáme voľbou ***Preprocess / Remove means ...***, ktorým nám vznikne dátový súbor s automaticky vygenerovaným názvom. Po dvojkliku na okienko daného setu môžeme zobraziť jeho podrobnosti a zmeniť jeho meno na *Norm. dáta*.

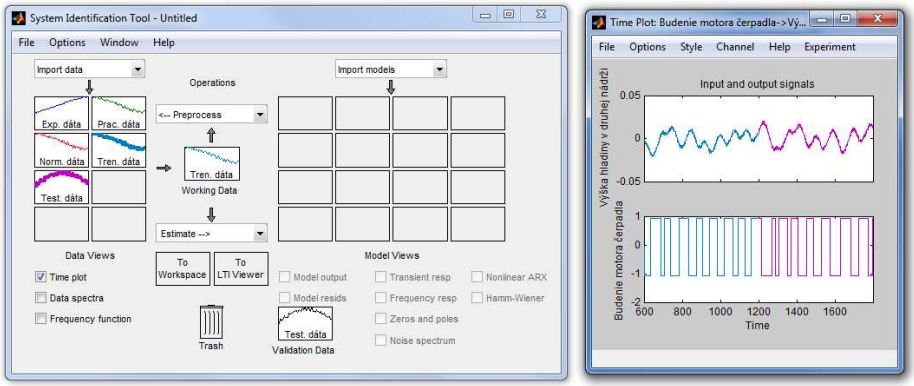


Obr. 8 Vľavo graf normovaných pracovných dát, v strede hlavné okno s dátovými setmi a vpravo okno s podrobnosťami dátového súboru Norm. dáta

C. Rozdelenie na trénovacie a testovacie dáta

V poslednom kroku predspracovania si dáta rozdelíme na polovicu. Na základe prvej polovice, ktorú nazveme *Tren. dáta* budeme vykonávať identifikáciu. Na základe druhej, ktorú nazveme *Test. data*, vykonáme výsledné porovnanie. Ako ***Working Data*** použijeme *Norm. dáta*. Rozdelenie je rovnaké, ako pri odstránení nábehu cez voľbu ***Preprocess / Select range ...***, rozdiel je iba v rozsahoch.

Po vytvorení je nutné nastaviť *Tren. dáta* ako ***Working Data*** a *Test. data* ako ***Validation data***. V tomto momente je možné vykonávať experimentálnu identifikáciu.

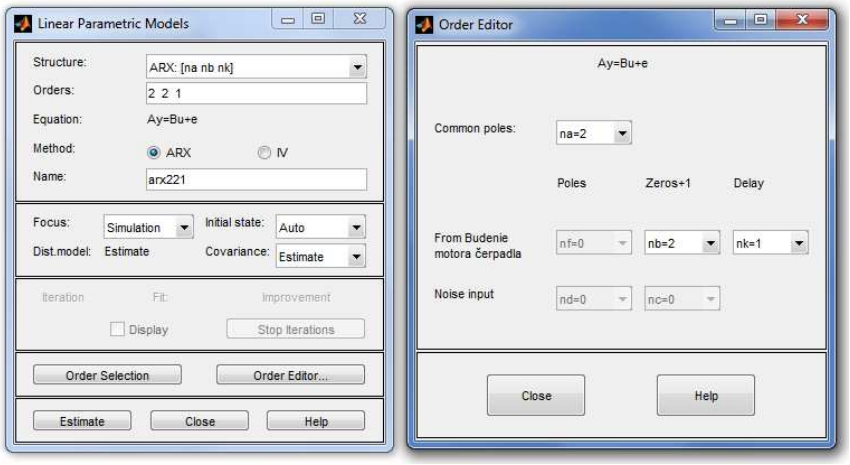


Obr. 9 Vľavo hlavné okno pripravené na experimentálnu identifikáciu, vpravo rozdelené dáta v IDPLOT-e

VI. EXPERIMENTÁLNA IDENTIFIKÁCIA NÁSTROJOM IDENT TOOL

Naším cieľom je nájsť aproximáciu modelom lineárnej dynamickej štruktúry. Tieto parametrické modely sú určené diskrétnym prenosom. Nás hlavne zaujímajú modely ARX a ARMAX [4].

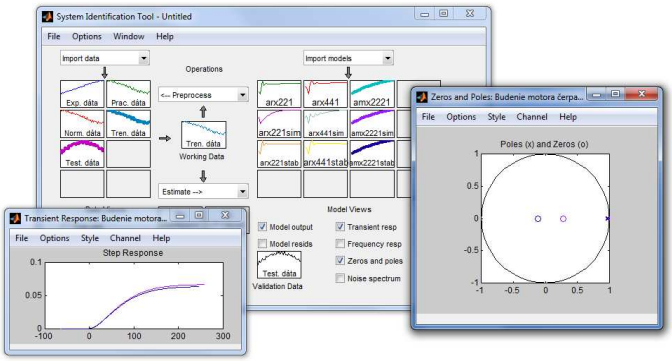
Cez voľbu ***Estimate / Linear parametric models ...*** môžeme určovať štruktúry modelov. Budeme pracovať iba s modelmi ARX a ARMAX. Rády je možné zadať ako vektor alebo tlačidlom ***Order editor*** otvoriť špecializované okno a navoliť podľa pólov, núl a filtra. Pod voľbou ***Focus*** vyberáme zameranie výpočtu. Prednastavenou hodnotou je ***Prediction*** – metóda vhodnejšia pre malé časové intervaly, zo štatistického hľadiska je optimálna. No môže viesť k nestabilnej aproximácii. Druhou možnosťou je ***Simulation*** – výhodná pre zašumené signály a pre modely bez filtra (ARX). Treťou možnosťou je **Stability** – rovnako ako *Prediction*, no výsledný model bude k stabilite donútený.



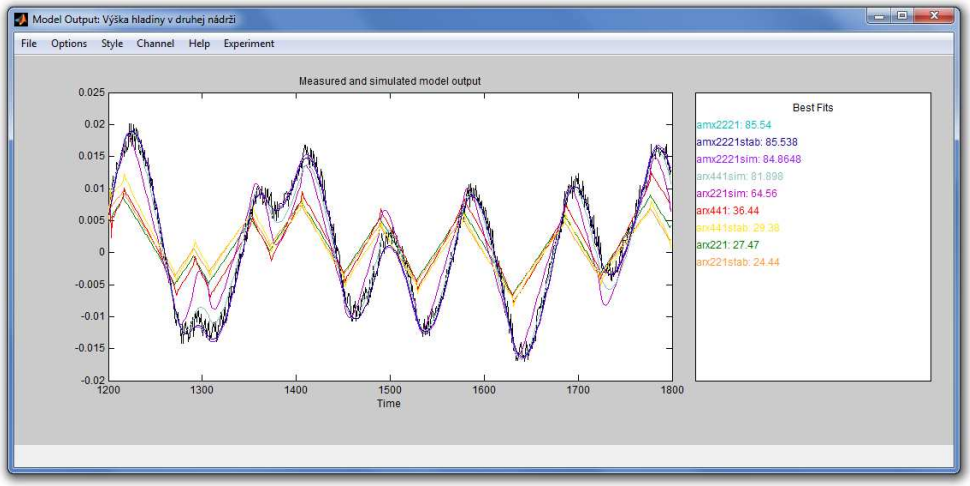
Obr. 10 Vľavo okno pre nastavenie odhadu parametrov modelom lineárnej štruktúry, vpravo okno editora rádov

Po potvrdení tlačidlom ***Estimate*** je vypočítaná aproximácia a pridaný model do hlavného okna IDENT tool. Výpočet aproximačných modelov vykonáme pre zameranie na *Predikciu*, *Simuláciu* a *Stabilitu* a to pre ARX221, ARX441 a ARMAX2221. Môžeme taktiež vykonať analýzu pólov a núl či odozvu na jednotkový skok pre zvolené modely. Výber týchto grafov sa nachádza v oblasti ***Model Views***.

Najdôležitejším krokom identifikácie je validácia výsledku. Spočíva v porovnaní skutočnej odozvy s aproximáciou, ktorá je budená rovnakým signálom. Tento graf s výpočtom zhody sa nachádza pod zaškrtávacou voľbou ***Transient resp***.



Obr. 11 Získané aproximačné modely v hlavnom okne, analýza ARMAX modelov v komplexnej rovine a odozva na jednotkový skok



Obr. 12 Získané aproximačné modely v hlavnom okne, analýza ARMAX modelov v komplexnej rovine a odozva na jednotkový skok

VII. ZÁVER

Z grafu porovnania a výpočtov zhody vyšli najlepšie modely ARMAX. Ostatné modely zlyhávali najmä kvôli relatívne veľkej perióde vzorkovania a šumu, ktorý nevedeli potlačiť. Pri zameraní za Simuláciu boli model ARX už pomerne presnejšie, no stále horšie ako ARMAX, pričom rád modelu na výsledok nemal vplyv.Taktiež je zrejmé, že vytvorenie podobných porovnaní pomocou bežných funkcii Matlab-u by bolo prácne a výsledný kód by bol málo neprehľadný. Preto na takto rozsiahle experimenty je IDENT tool veľmi výhodný

POĎAKOVANIE

Tento článok bol vytvorený realizáciou vedeckého projektu Vega č. 1/0286/11 Grantovej agentúry SR pod názvom Dynamické hybridné architektúry v multiagentových sieťových riadiacich systémoch (50%) a projektu Rozvoj Centra informačných a komunikačných technológií pre znalostné systémy (ITMS kód: 26220120030) na základe podpory operačného programu Výskum a vývoj financovaného z Európskeho fondu regionálneho rozvoja (50%).

LITERATÚRA

[1] ČERKALA, Jakub: Aplikácia výsledkov experimentálnej identifikácie v návrhu algoritmov riadenia reálneho modelu hydraulického systému. Diplomová práca. Košice: Technická univerzita v Košiciach, Fakulta elektrotechniky a informatiky, 2012. 83 s.
[2] ČERKALA, J. – JADLOVSKÁ A.: Aplikácia System Identification Toolboxu v experimentálnej identifikácii lineárnych dynamických systémov. Electrical Engineering and Informatics: Proceeding of the Faculty of Electrical Engineering and Informatics of the Technical University of Kosice. 2010, s. 540-545. ISSN 978-80-553-0460-1.
[3] LJUNG, L.: System Identification Toolbox, Používateľská príručka, 6. Verzia
[4] NOSKIEVIČ, P. : Modelování a identifikace systémů. Ostrava : Montanex a. s. 1999. 276 s. ISBN 80-7225-030-2

Ekonomické aspekty identifikácie malého prúdového motora

Marek ČERNEJ, Ladislav MADARÁSZ, Vladimír GAŠPAR

Katedra kybernetiky a umelej inteligencie, Fakulta elektrotechniky a informatiky,
Technická univerzita v Košiciach, Slovenská republika

marek.cernej@student.tuke.sk, ladislav.madarasz@tuke.sk, vladimir.gaspar@tuke.sk

Abstrakt — Tento článok poskytuje základnú analýzu experimentálnych dát získaných z Laboratória inteligentných systémov riadenia prúdových motorov. Úlohou je načrtnúť vzájomné pôsobenie merateľných činiteľov, ktoré majú podstatný vplyv na prevádzku malého prúdového motora MPM-20. Pomocou vybraného dataminingového nástroja sú rozobraté vzťahy medzi týmito činiteľmi a graficky znázornené vzájomné pôsobenie v konkrétnych fázach chodu motora.

***Kĺúčové slová* — data mining, malý prúdový motor, MPM-20, RapidMiner, experimentálna identifikácia**

I. ÚVOD

V dnešnej dobe sa v súvislosti s neustálym rozvojom prúdových leteckých motorov čoraz častejšie dostávajú do popredia otázky týkajúce sa ekonomických aspektov ich prevádzky. Otázky ako znížiť prevádzkové zaťaženie motora a s tým súvisiace opotrebenie, či ako znížiť spotrebu paliva alebo ako vôbec efektívne optimalizovať činitele s najväčším vplyvom na optimálny chod motora. Hľadanie odpovedí ne tieto otázky v praxi znamená vyhodnocovať množstvo historických dát z uskutočnených meraní, ich analýzu a v konečnom dôsledku prijímanie hodnotiacich záverov. Práve tu sa naskytá možnosť využitia sofistikovaných metód a nástrojov dataminingu pre efektívnu analýzu týchto dát.

Zámerom tohto článku je načrtnúť možnosti použitia konkrétnych dolovacích techník a nástrojov na historickej vzorke dát za účelom skúmania vplyvu rôznych merateľných veličín s podstatným vplyvom na prevádzku prúdových motorov.

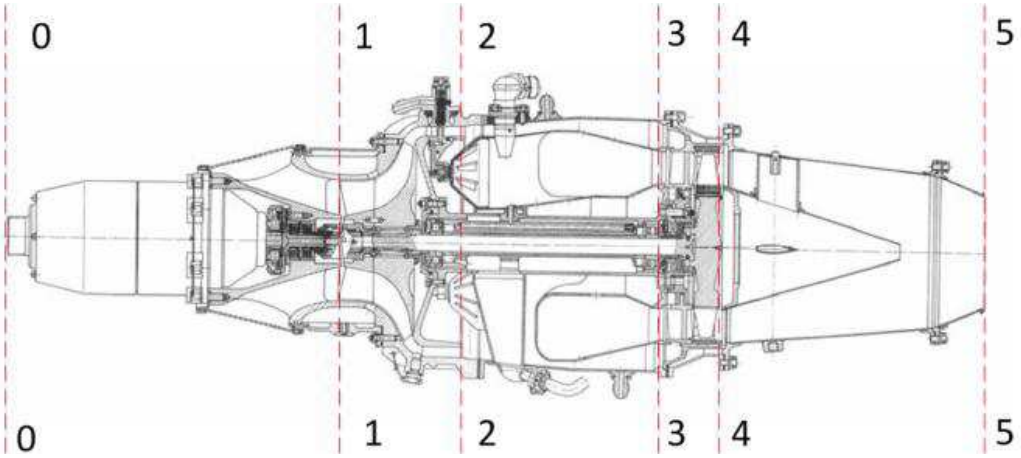
Článok obsahovo a konceptuálne vychádza najmä z diplomovej práce s názvom *Ekonomické aspekty prevádzky malého prúdového motora MPM-20 [7]*, kde je táto problematika rozoberaná podrobnejšie. Problematika výskumu ekonomických aspektov prevádzky MPM-20, jej aktuálnosť a potrebnosť je taktiež načrtnutá v monografii [1]. Ako základ pre analýzu boli použité experimentálne dáta získane z Laboratória inteligentných riadiacich systémov leteckých motorov (LIRSLM[1]) konkrétne z meraní prevádzky malého prúdového motora MPM-20.

II. PREDPOKLADY PRE ANALÝZU

A. Prúdový motor MPM-20

Zariadenie typu MPM-20 predstavuje letecký turbokompresorový spúšťač TS-21 sovietskej výroby využívaný na spúšťanie motorov AL-7F a AL21F. Pre potreby experimentálnej prevádzky bol upravený na jednoprúdový jednohriadeľový letecký motor s radiálnym kompresorom, združenou spaľovacou komorou, jednostupňovou nechladenou plynovou turbínou a pevnou výstupnou dýzou. Doba prevádzky tohto motora je teda obmedzená len na 1-2 minúty. Tento fakt budeme uplatňovať pri dekompozícii priebehu.

[1] LIRSLM - http://lirslm.fei.tuke.sk/



Obr. 1 Rezy motorom MPM-20 [5]

0 - rez pred motorom
1 - rez pred kompresorom
2 - rez za kompresorom
3 - rez za spaľovacou komorou
4 - rez za plynovou turbínou
5 - rez za výstupom motora (za výstupnou dýzou)

B. Prevádzkový režim

Priebeh meraní jednotlivých ukazovateľov je charakterizovaný chodom motora od jeho spustenia po dobeh. Štart motora začína spustením elektromotora, ktorý roztočí kompresor z 0 na približne 10 000 - 13 500 otáčok za minútu (*rpm*). Súčasne sa uvedie do prevádzky palivové čerpadlo, zapaľovacie sviečky a tiež olejové čerpadlo. Keď tlak na tlakovom spínači dosiahne hodnoty 0,02 – 0,045 kPa, elektromagnetický ventil sa otvorí a umožní tak vstrekovanie paliva a prietok oleja. Ak nastane zážih paliva a rýchlosť motora presiahne cca. 20 tis. otáčok za minútu, elektromotor sa spolu so zapaľovacími sviečkami odpája. Pri tlaku presahujúcom 0,19 – 0,21 kPa je tlakovým spínačom otvorený obtokový elektromagnetický ventil (EMV) označovaný ako bypass, ktorý reguluje prietok plynu v plynovej turbíne. Ideálny voľnobeh resp. tzv. kľudový stav dosahuje motor, ak je rýchlosť otáčok kompresora v rozmedzí cca. 40 tis. - 48 tis. otáčok za minútu. Pri tomto stave je príjem paliva stabilný, spaľovanie prebieha najmä s využitím vháňaného vzduchu. Po vypnutí motora sa automaticky znižuje tlak, uzavrie sa bypass, motor sa dostáva do fázy dobehu.

Z časového a prevádzkového pohľadu vieme priebeh teda rozdeliť na tri hlavné fázy [6]:
· fáza nábehu (otáčky pod 40 tis. v čase do 15 sekúnd)
· fáza voľnobehu (otáčky 40 tis. – 48 tis. v čase od 15 do 50 sekúnd)
· fáza dobehu (otáčky pod 40 tis. v čase nad 50 sekúnd)

C. Vzorka dát

Dáta, ktoré obsahuje analyzovaná vzorka, pochádzajú z prostredia LabView, ktoré zabezpečuje priamy zber týchto dát z jednotlivých senzorov. Nami použitá vzorka dát obsahuje údaje zo 41 meraní, integrované do jedného excelovského súboru, ktorý predstavuje základný zdrojový súbor našej analýzy. Časové ohraničenie vzorky je 09. 02. 2010 - 20. 05. 2010. Na samotné predspracovanie dát a dolovanie v nich sme použili softvér RapidMiner.

Sledovanými ukazovateľmi v rámci jednotlivých meraní sú:

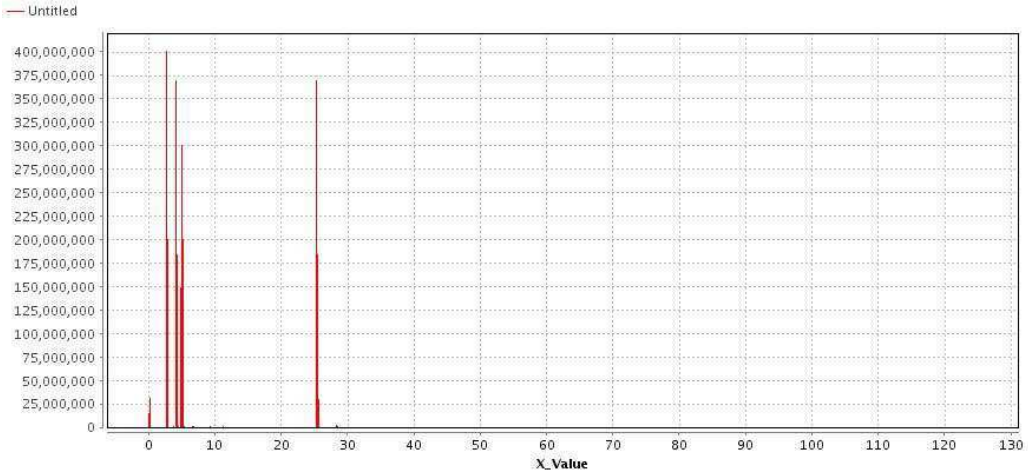
- **X_Value** – časový rozmer merania, mernou jednotkou je 0,1 sek., každá desatina sekundy determinuje jeden prípad, resp. vzor
- **T_{1c}, T_{2c}, T_{4c}** – ukazovatele teploty v rezoch prúdového motora, mernou jednotkou je °C
- **T_{ol}** – teplota oleja, mernou jednotkou je °C
- **Q_{pal}** – množstvo, resp. prietok paliva, mernou jednotkou je liter za minútu
- **P_{2c}, P_{3c}** – statický tlak v rôznych častiach motora, mernou jednotkou je atmosféra
- **P_{pal}** – tlak paliva, mernou jednotkou je atmosféra
- **Untitled** – otáčky turbíny resp. kompresora, mernou jednotkou je počet otáčok za minútu

III. ANALÝZA VZORKY DÁT

A. *Predspracovanie údajov*

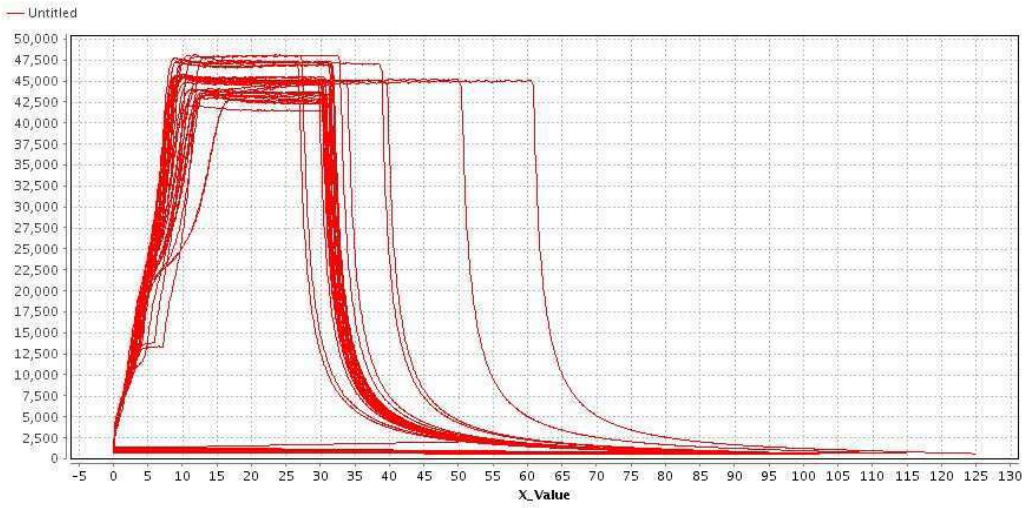
Kvalitné výsledky analýzy dosiahneme len na základe kvalitných údajov. Tie si zabezpečíme vhodným predspracovaním zdrojovej vzorky dát. Okrem už spomenutej integrácie dát je nevyhnutné odstrániť rôzne šumy, resp. chyby meraní spôsobených napr. chybou senzora. Konkrétne v našom prípade ide o senzor, ktorý meria otáčky motora. Pri predspracovaní dát bolo nevyhnutné vybrať len úspešné a normálne (s palivom) spustenia, nakoľko neúspešné merania menia charakter správania sa skúmaného objektu a tým pádom by skresľovali výsledky vo veľkej miere. Neúspešné merania je možné analyzovať samostatne pre hľadanie príčin vzniknutých porúch alebo problémov najmä vo fáze štartu a nábehu motora.

Zariadenie MPM-20 dosahuje ideálny chod (rýchlosť), ak sa otáčky kompresora pohybujú v rozmedzí 40 000 – 48 000 otáčok za minútu pri stabilnom príjme paliva čo v praxi znamená, že počet otáčok pri ideálnom chode motora nepresahuje úroveň 50 tis. otáčok za minútu. Avšak naša vzorka dát obsahuje hodnoty ukazovateľa otáčok (Untitled) vysoko nad 50 tis. ot/min, čiže je nevyhnutné takéto hodnoty odstrániť resp. odfiltrovať.



Obr. 2 Chybné hodnoty v meraniach otáčok (Untitled)

Po odfiltrovaní vychýlených hodnôt získavame stabilné a kvalitné dáta, na ktoré môžeme použiť zvolené techniky dolovania.



Obr. 3 Priebeh otáčok po odstránení chybných hodnôt

B. *Korelačná analýza*

Ako vhodnú metódu analýzy prevádzkových vlastností meraných ukazovateľov a ich vplyv na chod prúdového motora sme si zvolil korelačnú analýzu. Cieľom je zistiť silu vzájomného pôsobenia jednotlivých meraných atribútov. Dôležitým ukazovateľom teda bude korelačný

koeficient, ktorý nám túto vzájomnú závislosť dokáže interpretovať.

Pre výpočet korelačného koeficientu platí vzťah:

$$r_{xy} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sqrt{\overline{x^2} - \left(\bar{x}\right)^2} \cdot \sqrt{\overline{y^2} - \left(\bar{y}\right)^2}} \tag{1}$$

kde r_{xy} je označenie korelačného koeficientu, $\bar{x}$ a $\bar{y}$ sú výberové priemery náhodných veličín x, y a výrazy $\overline{x^2} - \left(\bar{x}\right)^2$ resp. $\overline{y^2} - \left(\bar{y}\right)^2$ sú modifikované výberové smerodajné odchýlky.

Výsledkom použitia korelačnej analýzy na celkový priebeh prevádzky motora je tabuľka korelačných koeficientov. Na jej základe vieme pre každý z atribútov konštatovať určitú závislosť voči ostatným meraným atribútom (veličinám). Na meranie lineárnej závislosti pritom existuje orientačná stupnica, ktorá je definovaná nasledovne:

- Slabá závislosť, ak $0 < |r| \leq 0{,}3$
- Mierna resp. stredná závislosť, ak $0{,}3 < |r| \leq 0{,}8$
- Silná závislosť, ak $0{,}8 < |r| \leq 1$

C. Korelačné vzťahy z pohľadu prevádzkových intervalov

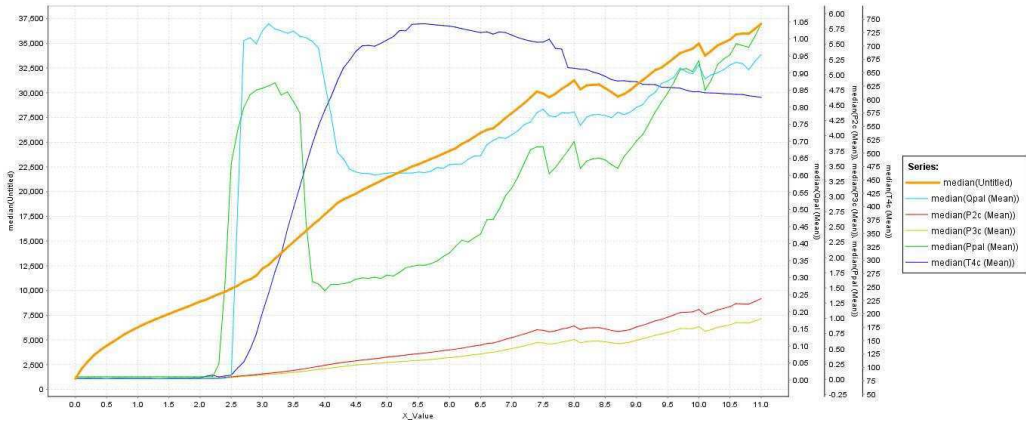
Korelačnú analýzu v prostredí RapidMiner sme z dôvodu vyššej výpovednej hodnoty aplikovali na jednotlivé prevádzkové intervaly definované v predchádzajúcej kapitole. Je to najmä z toho dôvodu, že jednotlivé merané ukazovatele dosahujú v konkrétnych časových intervaloch chodu motora odlišné hodnoty, ktoré môžu významne skresliť celkovú výpovednú hodnotu analýzy.

Fáza rozbehu, resp. nábehu prebieha od štartu elektromotora po nábeh do voľnobehu. V prostredí RapidMiner vieme tento interval extrahovať napr. pomocou filtrovania hodnôt celkového chodu. Dôležité je najmä stanoviť ideálnu približnú hranicu otáčok, kedy motor z fázy nábehu prejde do fázy voľnobehu. V našom prípade sme si zvoli približnú hranicu 41 300 otáčok za minútu [6]. Vzhľadom na to, že otáčky sa pod hranicu 41 300 rpm dostávajú aj vo fáze dobehu, sme interval nábehu motora ohraničili časove na 11 sekúnd. Po definovaní intervalu a aplikovaní operátora korelačnej analýzy dostávame tabuľku korelačných koeficientov pre fázu nábehu (Obr. 4). Na základe nej vieme povedať, že na nábehovú fázu motora majú z meraných atribútov podstatný vplyv najmä otáčky (Untitled), teplota T_{4c}, prietok a tlak paliva, a statické atribúty tlaku P_{2c}, P_{3c}.

Attributes	X_Value	T1c (Mean)	T2c (Mean)	T4c (Mean)	Tol (Mean)	Qpal (Mean)	P2c (Mean)	P3c (Mean)	Ppal (Mean)	Untitled
X_Value	1	0.319	0.327	0.525	0.355	0.358	0.716	0.706	0.421	0.849
T1c (Mean)	0.319	1	0.018	0.213	0.037	0.227	0.291	0.288	0.215	0.354
T2c (Mean)	0.327	0.018	1	0.174	0.357	0.056	0.499	0.498	0.228	0.380
T4c (Mean)	0.525	0.213	0.174	1	0.326	0.276	0.475	0.471	0.172	0.708
Tol (Mean)	0.355	0.037	0.357	0.326	1	0.065	0.300	0.290	0.092	0.378
Qpal (Mean)	0.358	0.227	0.056	0.276	0.065	1	0.289	0.287	0.636	0.399
P2c (Mean)	0.716	0.291	0.499	0.475	0.300	0.289	1	1.000	0.557	0.895
P3c (Mean)	0.706	0.288	0.498	0.471	0.290	0.287	1.000	1	0.558	0.889
Ppal (Mean)	0.421	0.215	0.228	0.172	0.092	0.636	0.557	0.558	1	0.514
Untitled	0.849	0.354	0.380	0.708	0.378	0.399	0.895	0.889	0.514	1

Obr. 4 Tabuľka korelačných koeficientov pre nábeh motora

V praxi to znamená, že sa najprv spustí elektromotor, ktorým sa dosiahne prevádzková rýchlosť do približne 13 500 otáčok za minútu. Súčasne sa spustia zapaľovacie sviečky, palivové čerpadlo a olejové čerpadlo. Samotné nasatie paliva je možné vidieť aj na grafe v podobe prudšieho počiatočného výkyvu, ktorý zaznamenali senzory prietoku paliva a tlaku paliva. Počas nábehu kompresora sa zvyšuje tlak P_{2c} a P_{3c}. Keď tlak dosahuje úroveň približne 0,02 až 0,045 kPa, otvorí sa elektromagnetický ventil a umožní tak prietok paliva (Q_{pal}) a jeho vstrekovanie do spaľovacej komory. Vstrekovanie paliva je priamo úmerné jeho tlaku (P_{pal}). Súčasne je otvorený prietok oleja. Teplotný senzor T_{4c} indikuje teplotu na výstupe z plynovej turbíny, ktorá premieňa spaliny zo spaľovacej komory na kinetickú energiu, pomocou ktorej roztáča lopatky turbíny. Pri nábehu motora teplota T_{4c} prudko stúpne a pohybuje na úrovni 700 až 800 ºC až kým otáčky nedosiahnu približne 40 tis. rpm. Vtedy motor prechádza do fázy voľnobehu.



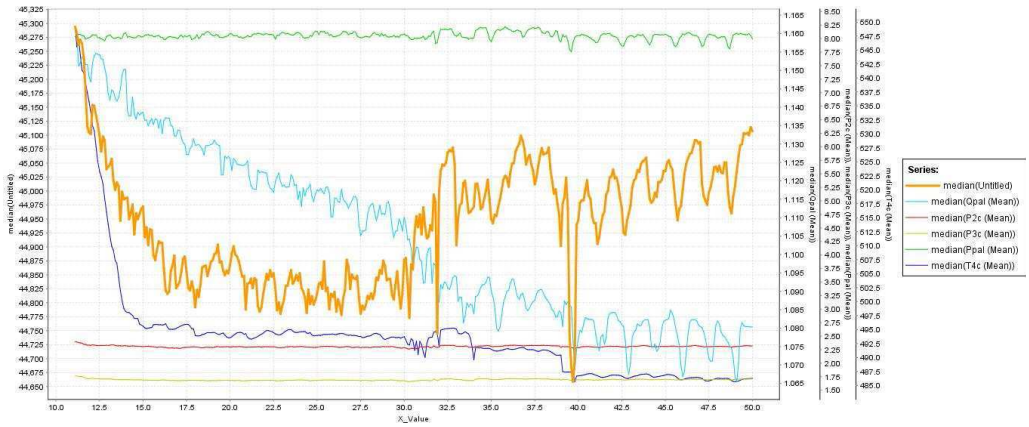
Obr. 5 Nábeh motora (T_{4c}, Q_{pal}, P_{pal}, P_{2c}, P_{3c}, Untitled)

Fáza voľnobehu začína nad úrovňou 40 tis. otáčok za minútu. Podobne ako pri fáze nábehu sme v prostredí RapidMiner tento interval extrahovali pomocou filtrovania otáčok a časového priebehu meraní. V prípade otáčok sme si stanovili hranice od 41 300 do 50 000 rpm. Pri časovom rozhraní sme stanovili rozmedzie 12 až 50 sekúnd. Po použití operátora korelačnej analýzy dostávame tabuľku korelačných koeficientov pre fázu voľnobehu (Obr. 6). Na základe nej môžeme konštatovať, že najsilnejšiu vzájomnú koreláciu sledujeme najmä medzi atribútmi tlaku a spotreby paliva voči otáčkam motora, kedy prevádzka motora funguje už iba na samospaľovacom princípe.

Attributes	X_Value	T1c (Mean)	T2c (Mean)	T4c (Mean)	Tol (Mean)	Qpal (Mean)	P2c (Mean)	P3c (Mean)	Ppal (Mean)	Untitled
X_Value	1	0.107	0.241	0.083	0.193	0.106	0.000	0.001	0.007	0.000
T1c (Mean)	0.107	1	0.018	0.032	0.000	0.320	0.441	0.460	0.300	0.368
T2c (Mean)	0.241	0.018	1	0.072	0.084	0.046	0.374	0.339	0.171	0.378
T4c (Mean)	0.083	0.032	0.072	1	0.042	0.159	0.019	0.030	0.078	0.006
Tol (Mean)	0.193	0.000	0.084	0.042	1	0.040	0.030	0.023	0.022	0.043
Qpal (Mean)	0.106	0.320	0.046	0.159	0.040	1	0.495	0.522	0.651	0.448
P2c (Mean)	0.000	0.441	0.374	0.019	0.030	0.495	1	0.989	0.644	0.981
P3c (Mean)	0.001	0.460	0.339	0.030	0.023	0.522	0.989	1	0.701	0.971
Ppal (Mean)	0.007	0.300	0.171	0.078	0.022	0.651	0.644	0.701	1	0.624
Untitled	0.000	0.368	0.378	0.006	0.043	0.448	0.981	0.971	0.624	1

Obr. 6 Tabuľka korelačných koeficientov vo fáze voľnobehu

Motor dosahuje kľudový stav (voľnobeh) pri otáčkach kompresora od 40 tis. - 48 tis. otáčok za minútu. Príjem paliva je udržiavaný stabilne na úrovni 1,1 litra za minútu, pričom jeho tlak postupne klesá pod úroveň 3 atmosfér. Ostatné tlakové senzory sú ustálené v rozmedzí 1,5 až 2,5 atmosfér. Teplota T_{4c} postupne klesá na úroveň približne 450 ºC. Hodnota otáčok je priamo závislá na spotrebe paliva, ktoré sa zmiešava so vzduchom v spaľovacej komore a vytvára zápalnú zmes. Prúdom spalín sú poháňané lopatky turbíny.

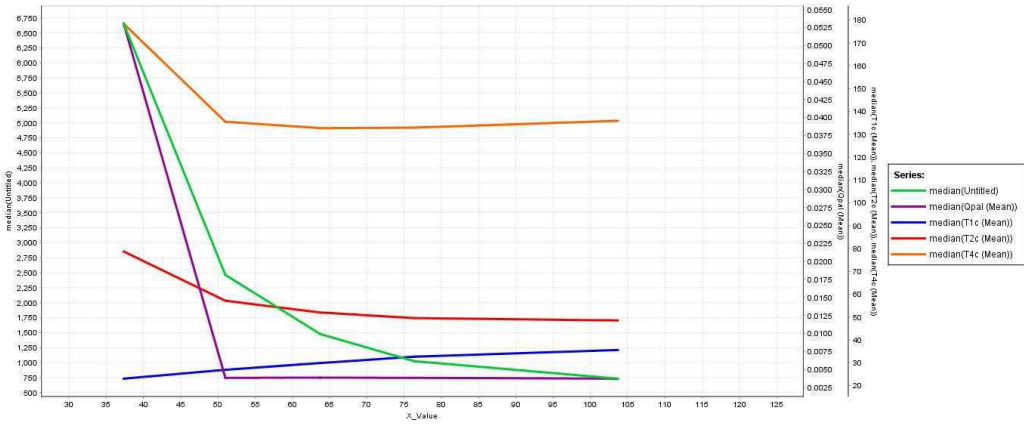


Obr. 7 Voľnobeh motora (T_{4c}, Q_{pal}, P_{pal}, P_{2c}, P_{3c}, Untitled)

Po vypnutí motora sa exponenciálne zníži tlak vzduchu až na nulovú hranicu, uzavrie sa prítok paliva, motor sa dostáva do fázy dobehu, otáčky sa znižujú k nulovej hodnote, motor chladne. Podobne ako pri fáze nábehu alebo voľnobehu sme v prostredí RapidMiner tento interval extrahovali pomocou filtrovania otáčok a časového priebehu meraní. V prípade otáčok sme si stanovili hranicu pod 41 300 *rpm*. Pri časovom rozhraní sme stanovili rozmedzie 30 až 180 sekúnd.

Attributes	X_Value	T1c (Mean)	T2c (Mean)	T4c (Mean)	Tol (Mean)	Qpal (Mean)	P2c (Mean)	P3c (Mean)	Ppal (Mean)	Untitled
X_Value	1	0.724	0.449	0.242	0.295	0.128	0.076	0.074	0.022	0.355
T1c (Mean)	0.724	1	0.495	0.292	0.090	0.182	0.125	0.122	0.037	0.405
T2c (Mean)	0.449	0.495	1	0.860	0.219	0.594	0.473	0.464	0.149	0.899
T4c (Mean)	0.242	0.292	0.860	1	0.168	0.813	0.716	0.713	0.298	0.960
Tol (Mean)	0.295	0.090	0.219	0.168	1	0.067	0.040	0.037	0.009	0.200
Qpal (Mean)	0.128	0.182	0.594	0.813	0.067	1	0.950	0.939	0.498	0.824
P2c (Mean)	0.076	0.125	0.473	0.716	0.040	0.950	1	0.994	0.577	0.727
P3c (Mean)	0.074	0.122	0.464	0.713	0.037	0.939	0.994	1	0.641	0.719
Ppal (Mean)	0.022	0.037	0.149	0.298	0.009	0.498	0.577	0.641	1	0.300
Untitled	0.355	0.405	0.899	0.960	0.200	0.824	0.727	0.719	0.300	1

Obr. 8 Tabuľka korelačných koeficientov vo fáze dobehu



Obr. 9 Dobeh resp. dojazd motora (T1c, T2c, T4c, Qpal, Untitled)

IV. ZHODNOTENIE

Aj vďaka tejto analýze a jej výsledkom si dokážeme lepšie predstaviť a zhodnotiť vzájomný vplyv a pôsobenie jednotlivých meraných činiteľov. To, ako na seba jednotlivé činitele vplývajú, nám umožňuje navrhnúť optimálne prevádzkové podmienky pre efektívny priebeh chodu prúdového motora od jeho spustenia cez voľnobeh až po jeho dobeh. S ohľadom na rozsah tohto článku nebolo možné zodpovedať všetky otázky, ktoré sa nám núkajú. Môžeme len predostrieť odporúčania pre ďalšie analýzy a výskum.

Jednou z možných výziev je napríklad vytvorenie predikčného modelu, na základe ktorého dokážeme s ohľadom na historické dáta predikovať zmeny a simulovať správanie sa jednotlivých atribútov pri zmene konkrétnych hodnôt bez nevyhnutnosti reálneho experimentu.

V. ZÁVER

Ako už v úvode tohto článku bolo naznačené, samotná oblasť skúmania je vysoko aktuálna, keďže s neutíchajúcim rozmachom avionických systémov je potrebné sa touto problematikou neustále zaoberať a hľadať nové možnosti zlepšovania funkčných vlastností týchto zložitých systémov. V tomto objavovaní nám pomáhajú sofistikované metódy, techniky a softvérové nástroje, ktoré predstavujú v oblasti dataminingu vrchol vedeckého bádania a aj vďaka tomu môžeme napredovať na ceste poznania a zlepšovania poznaného.

POĎAKOVANIE

The work presented in this paper was supported by VEGA, Grant Agency of Ministry of Education and Academy of Science of Slovak Republic under Grant No. 1/0298/12 – “Digital control of complex systems with two degrees of freedom”. The work presented in this paper was

also supported by KEGA under Grant No. 018TUKE-4/2012 - Progressive methods of education in the area of control and modeling of complex systems object oriented on aircraft turbo-compressor engines. This support is very gratefully acknowledged.

LITERATÚRA

[1] T. Lazár, L. Madarász (Eds.): Inovative Outputs from the Transformed Experimental Laboratory with a Small Turbojet Engine. elfa, s.r.o. Košice, 348pp. ISBN 978-80-8086-170-4 (2011).

[2] L. Madarász, R. Andoga, L. Főző: Intelligent Technologies in Modeling and Control of Turbojet Engines. In: New Trends in Technologies: Control, Management, Computational Intelligence and Network Systems, Meng Joo Er (Ed.), Sciyo, Available on: http://www.intechopen.com/articles/show/title/intelligent-technologies-in-modeling-and-control-of-turbojet-engines, pp. 17 -38. ISBN: 978-953-307-213-5, (2010).

[3] T. Karoľ, V. Gašpar, L. Madarász, R. Andoga, L. Főző, J. Judičák: Advanced approaches Artificial intelligence in diagnostics of the turbojet engine MPM-20, 12th IEEE International Symposium on Computational Intelligence and Informatics, 21. - 22. November, Budapest, Hungary, ISBN 978-1-4577-0043-9, pp. 259 – 264, (2011).

[4] L. Madarász, T. Karoľ, V. Gašpar, R. Andoga, L. Főző and J. Judičák: Digital start-up control of the small turbojet engine MPM-20. 5pp. In: proceedings SAMI 2012 Herľany, Slovak Republic. ISBN 978-1-4577-0195-5 (2012).

[5] V. Gašpar, L. Madarász, R. Andoga et al.: Replacement impacts of MPM-20 turbojet engine using measured data analysis. In ELITECH ’12 – 14th conference of Doctoral Students: proceddings. Bratislava, 2012 (accepted).

[6] V. Gašpar, T. Karoľ, C. Havrilová: Ideal MPM-20 jet engine running process with zero degrees of freedom. In SCYR 2012 – 12th Scientific Conference of Young Researchers: proceddings. FEI TU of Košice, 2012.

[7] M. Černej: Ekonomické aspekty prevádzky malého prúdového motora MPM-20, Diplomová práca. FEI TU, Košice, 2012

Analýza dosiahnuteľnosti stavu výrobných liniek použitím Petriho sieti

Matej ČOPÍK

Katedra kybernetiky a umelej inteligencie, Fakulta elektrotechniky a informatiky,
Technická univerzita v Košiciach, Slovenská republika

matej.copik@tuke.sk

***Abstrakt* — Tento článok sa venuje analýze dosiahnuteľnosti stavu výrobných liniek, ktoré sú namodelované pomocou Petriho sieti. Uvedená metodika je použiteľná aj na iné Petriho siete, ktoré spĺňajú požadované vlastnosti. Dosiahnuteľnosť stavu predstavuje možnosť transformovať Petriho sieť z počiatočného do požadovaného stavu. Dosiahnuteľnosť v prípade výrobných liniek znamená možnosť zastavenia alebo prerušenia výrobného procesu. Analýzou dosiahnuteľných stavov je možné určiť vlastnosti výrobnej linky a vypočítaná prepaľovacia sekvencia sa dá použiť na ďalšiu analýzu času výroby alebo nákladov.**

***Kľúčové slová* — Petriho sieť, výrobná linka, incidenčná matica, prepaľovacia sekvencia**

I. ÚVOD

Výrobné systémy prestavujú budúcnosť priemyslu. V súčasnosti je veľmi rozšírené automatizovať výrobu plno alebo polo automatizovanými výrobnými linkami. Výrobná linka môže predstavovať časť výrobného systému alebo procesu. Súčasným trendom je budovať závody budúcnosti, v ktorých sú všetky vykonávané činnosti automatizované. Niektoré závody však venujú pozornosť aj modernizácii výrobných liniek, v ktorých boli jednotlivé operácie vykonávane manuálne. Automatizáciou takýchto výrobných liniek sa zvyšuje efektivita a kvalita vyrábaných výrobkov. Pri automatizovaných výrobných linkách je možné vykonávať aj náročné operácie, ktoré človek nezvládal. Preto vznikajú a zanikajú niektoré stavy výrobného procesu. Identifikácia takýchto stavov je dôležitá hlavne z pohľadu optimalizácie výroby. Špecifikácia a analýza stavov výrobných liniek je výhodná už počas projektovania, či sa jedná o nové alebo o upravované výrobné linky. Výrobné linky je možné už v projektovej fáze modelovať a analyzovať. Jedným z nástrojov na modelovania a analyzovanie výrobných liniek sú aj Petriho siete (PS). Pomocou modelu je možné analyzovať dosiahnuteľnosť stavov výrobných liniek a následne ich analyzovať. Analýza výrobnej linky v projektovej fáze je veľmi vhodná, pretože prípadne úpravy sú jednoducho zapracované do projektu. Upravená výrobná linka sa opäť namodeluje pomocou PS a následne analyzuje. Po získaní požadovaných výsledkov z modelu je možné výrobnú linku zostrojiť. Reálna výrobná linka bude mať vlastnosti modelu, ktorý bol analyzovaný už v projektovej fáze.

II. DOSIAHNUTEĽNOSŤ STAVU PETRIHO SIETI

Dosiahnuteľnosť v PS je náročný problém a má významne postavenie z hľadiska analýzy vlastnosti ako živosť, ohraničenie, pokrytie, zablokovanie a iné. Tieto a ďalšie vlastnosti je možné nájsť v [7]. Problém sa zväčšuje pri časovaných PS. Pri časovaných PS nie je dôležitá iba dosiahnuteľnosť, ale aj dosiahnuteľnosť v stanovenom čase. Jedná z metód riešenia problému dosiahnuteľnosti je prezentovaná v [5], ktorá pozostáva z transformácie problému dosiahnuteľnosti v PS na problém celočíselného lineárneho programovania. Ďalšie metódy, ktoré sa snažia riešiť problém dosiahnuteľnosti, sú založené na dekompozičnom prístupe. V [4] sú navrhnuté tri spôsoby na riešenie problému dosiahnuteľnosti pomocou metódy dekompozície. Spôsob popísaný v tomto článku je založený na výpočte dosiahnuteľnosti modifikáciou stavovej rovnice popisujúcej PS. Pre výpočet dosiahnuteľnosti touto metódou musí PS spĺňať niektoré vlastnosti ako živosť, bezkonfliktnosť a nesmie obsahovať vlastné cykly. Stavová rovnica matematicky popisuje zmeny v značení v PS. Matica $\boldsymbol{M_0}$ predstavuje počiatočné značenie PS a matica $\boldsymbol{M_k}$ predstavuje značenie po k-tom prepálení prechodu (prechodov). V ďalšom texte bude používaný výraz po k-tom kroku PS. Obe matice majú

rozmer m x 1, čiže sú to stĺpcové vektory obsahujúce aktuálny počet značiek v jednotlivých miestach. Riadiaci vektor $\boldsymbol{u_k}$ určuje, ktorý prechod sa prepálil v k-tom kroku. Vektor $\boldsymbol{u_k}$ je stĺpcový vektor n x 1, ktorý obsahuje n-1 núl a jednu jednotku, ktorá určuje prepálený prechod. Matica $\boldsymbol{C}$ je incidenčná matica, ktorá popisuje PS. Pre PS je teda možné napísať stavovú rovnicu v tvare:

$$\boldsymbol{M_k} = \boldsymbol{M_{k-1}} + \boldsymbol{C} * \boldsymbol{u_k}, \qquad kde\ k = 1, 2, 3, \ldots . \quad (1)$$

Za predpokladu, že želané značenie $\boldsymbol{M_d}$ je dosiahnuteľné z počiatočného značenia $\boldsymbol{M_0}$ pomocou postupnosti prepálení ($\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_d}$). Sumáciu stavových rovníc (1) pre $i = 1, 2, \ldots, d$ je možné zapísať v tvare:

$$\boldsymbol{M}_d = \boldsymbol{M_0} + \boldsymbol{C} * \sum_{k=1}^{d} \boldsymbol{u_k}. \quad (21)$$

Zápis (2) je možné upraviť do tvaru:

$$\boldsymbol{M}_d = \boldsymbol{M_0} + \boldsymbol{C} * \sigma, \quad (3)$$

kde $\boldsymbol{\sigma} = \sum_{k\ -\ 1}^{d} \boldsymbol{u_k}$.

Vektor $\boldsymbol{\sigma}$ je stĺpcový vektor obsahujúci nezáporné celé čísla rozmeru n x 1 a nazýva sa vektor počtu prepálení alebo prepaľovacia sekvencia. Jednotlivé zložky vektora $\boldsymbol{\sigma}$ označujú koľkokrát sa príslušný prechod musí prepáliť, aby došlo k transformácii PS zo značenia $\boldsymbol{M_0}$ do značenia $\boldsymbol{M_d}$ [6].

Rovnica (3) vypočítava dosiahnuteľný stav pri známej prepaľovacej sekvencií. Pri analýze dosiahnuteľnosti je však potrebné analyzovať, či vybraný stav je alebo nie je dosiahnuteľný.

Definícia 1

Stav $\boldsymbol{M_x}$ v PS je dosiahnuteľný ak existuje prepaľovacia sekvencia, ktorá transformuje PS z počiatočného stavu $\boldsymbol{M_0}$ do želaného stavu $\boldsymbol{M_x}$. Prepaľovacia sekvencia, musí nadobúdať hodnoty celých nezáporných čísel.

Pre analýzu dosiahnuteľnosti stavu $\boldsymbol{M_x}$ je však potrebné rovnicu (3) modifikovať do tvaru:

$$\boldsymbol{\sigma} = \boldsymbol{C}^{+} * (\boldsymbol{M_x} - \boldsymbol{M_0}), \quad (4)$$

kde $\boldsymbol{C}^{+}$ je pseudoinverzná matica k incidenčnej matici $\boldsymbol{C}$.

Pseudoinverzná matica je použitá, pretože k mactici $\boldsymbol{C}$ neexistuje inverzná matica. Incidenčá matica $\boldsymbol{C}$ je neštvorcová a vo väčšine prípadov riedka. Pri výpočte pseudoinverznej matice bol použitý Penroseov algoritmus s pevne danou ukončovacou podmienkou. Algoritmus na výpočet pseudoinverznej matice je možné nájsť v [1].

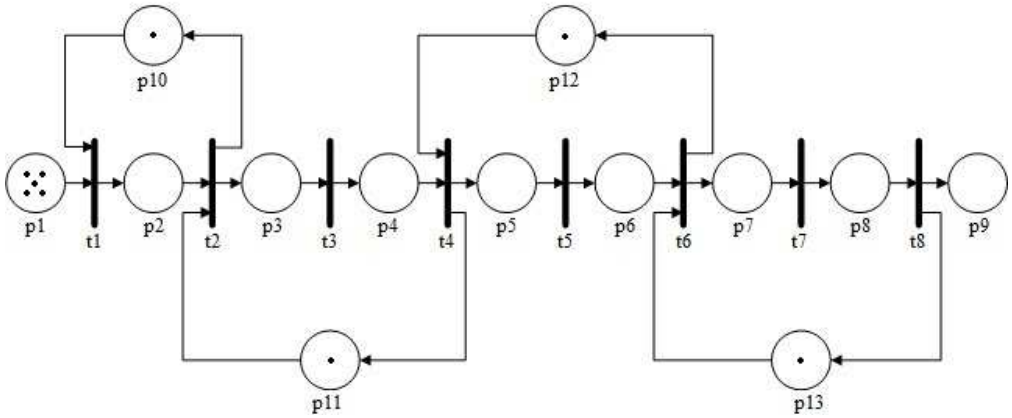
III. MODEL VÝROBNEJ LINKY

Za výrobnú linku možno považovať aj výučbový model výrobnej linky, ktorý sa nachádza na Katedre kybernetiky a umelej inteligencie (KKUI) Technickej univerzity (TU) v Košiciach v laboratóriu V147. Tento model nie je určený pre priemyselnú výrobu, ale slúži pre výuku a výskum. Model predstavuje plne automatizovanú výrobnú linku. Popis výrobnej linky je možné nájsť v [2] a [3]. Na obrázku obr.1 je zobrazená PS, ktorá modeluje danú výrobnú linku. Model je zjednodušený pre lepšie pochopenie uvedenej metodiky. Zjednodušenie modelu nemá vplyv na analýzu dosiahnuteľnosti a uvedený postup je možné použiť aj na zložité modely. PS, ktorú chceme analyzovať, musí spĺňať niektoré vlastnosti ako dosiahnuteľnosť, živosť, bezkonfliktnosť a nesmie obsahovať vlastné cykly.

Popis jednotlivých miest a prechodov je nasledovný:

p1 – paleta je na vstupnom sklade postu 3 a čaká na uvoľnenie postu 4,
t1 – presun palety na post 4,
p2 – paleta na poste 4 je pripravená na presun na druhý dopravník, čaká na uvoľnenie postu 5,
t2 – presun palety z dopravníka 1 na dopravník 2 a následne na post 5, uvoľnenie postu 4,
p3 – plná paleta na poste 5 je pripravená na vyprázdnenie,
t3 – vyprázdnenie palety na poste 5,
p4 – prázdna paleta na poste 5 je pripravená na presun a čaká na uvoľnenie postu 1,
t4 – presun palety na post 1, uvoľnenie postu 5,
p5 – prázdna paleta je na poste 1 a čaká na vyskladanie obrazca,
t5 – vyskladávanie obrazca do palety na poste 1,
p6 – plná paleta na poste 1 je pripravená na presun a čaká na uvoľnenie postu 2,
t6 – presun palety na post 2, uvoľnenie postu 1,

p7 - paleta na poste 2 čaká na koniec kamerovej kontroly,
t7 – kamerová kontrola palety na poste 2,
p8 - paleta po kamerovej kontrole na poste 2 je pripravená na presun na výstupný sklad postu 3,
t8 – presun palety na výstupný sklad postu 3, uvoľnenie postu 2,
p9 – paleta je na výstupnom sklade postu 3,
p10 – post 4 je voľný,
p11 – post 5 je voľný,
p12 – post 1 je voľný,
p13 – post 2 je voľný.



Obr. 1 model výrobnej linky

PS z obrázka obr.1 je namodelovaná ako zovšeobecnená PS a je ju možné popísať pomocou usporiadanej pätice:

$$N = (\boldsymbol{P}, \boldsymbol{T}, \boldsymbol{pre}, \boldsymbol{post}, \boldsymbol{M_0}) \tag{5}$$

kde $\boldsymbol{P}$ = {p1, p2, ..., pm} je konečná množina miest,
$\boldsymbol{T}$ = {t1, t2, …, tn} je konečná množina prechodov,
pre - priama incidenčná funkcia, ***pre***: $\boldsymbol{P} \times \boldsymbol{T} \rightarrow N$,
post - spätná incidenčná funkcia, ***post***: $\boldsymbol{P} \times \boldsymbol{T} \rightarrow N$,
$\boldsymbol{M_0}$ - začiatočné označenie $\boldsymbol{M_0} : \boldsymbol{P} \rightarrow N$,
N - množina celých nezáporných čísel

Z modelu PS je pre analýzu dosiahnuteľnosti poznať počiatočné značenie $\boldsymbol{M_0}$ a incidenčnú maticu $\boldsymbol{C}$, ktoré sú pre PS z obrázka obr.1 zobrazené nižšie.

$$\boldsymbol{M}_0 = \begin{bmatrix} 5 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \qquad \boldsymbol{C} = \begin{bmatrix} -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 \end{bmatrix} \tag{6}$$

IV. Analýza dosiahnuteľnosti stavu v Petriho sieti

Dosiahnuteľnosť stavu PS je možné vypočítať zo vzťahu (4). Incidenčná matica a počiatočný stav je známy z popisu PS. Požadovaný stav výrobnej linky je potrebné určiť podľa toho aký reálny stav výrobnej linky sa má dosiahnuť. Ako prvý stav, ktorého dosiahnuteľnosť je potrebné určiť, je stav $\boldsymbol{M_x}$ kedy je prvá paleta vyskladaná a na výstupnom sklade ešte nie je výrobok.

Požadovaný stav $\boldsymbol{M_x}$ je možná zapísať v podobe vektora (7).

$$\boldsymbol{M}_x = \begin{bmatrix} 4 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \tag{7}$$

Želaný stav sa dosadí do rovnice (4) a vypočíta sa prepaľovacia sekvencia.

$$\boldsymbol{\sigma} = \boldsymbol{C}^{+} * \left(\begin{bmatrix} 4 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 5 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \right) = \begin{bmatrix} 0,85 \\ 0,78 \\ 0,73 \\ 0,68 \\ 0,96 \\ 0,25 \\ 0,20 \\ 0,15 \end{bmatrix} \tag{8}$$

Vypočítaná prepaľovacia sekvencia je neuskutočniteľná, z čoho vyplýva že požadovaný stav nie je dosiahnuteľný. Prepaľovacia sekvencia, musí nadobúdať hodnoty celých nezáporných čísel, čiže prechod sa buď vykoná alebo nie.

Prepaľovacia sekvencia je neuskutočniteľná pretože ak je paleta vyskladaná a čaká na presun na poste 1, tak post 1 nie je voľný. Stav $\boldsymbol{M_x}$ uvažoval, že post 1 je voľný, preto je potrebné nový požadovaný stav $\boldsymbol{M_y}$ modifikovať. Stav $\boldsymbol{M_y}$ hovorí to isté čo stav $\boldsymbol{M_x}$, ale s tým rozdielom že post 1 je obsadený. Rovnica (9) vypočítava prepaľovaciu sekvenciu pre nový želaný stav $\boldsymbol{M_y}$.

$$\boldsymbol{\sigma} = \boldsymbol{C}^{+} * \left(\begin{bmatrix} 4 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \\ 1 \end{bmatrix} - \begin{bmatrix} 5 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \right) = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{9}$$

Vypočítaná prepaľovacia sekvencia je uskutočniteľná, z čoho vyplýva že požadovaný stav je dosiahnuteľný. Obdobným spôsobom je možné určiť dosiahnuteľnosť ľubovoľného stavu.

V. ZÁVER

Súčasným trendom v budovaní výrobných liniek je tvorba závodov budúcnosti, ktoré tvoria plno automatizované prvky. Spájaním jednotlivých výrobných liniek do výrobných celkov sa v niektorých prípadoch mení spôsob a účel použitia jednotlivých častí. Samotný návrh takýchto zoskupení je veľmi náročný a po realizácii často dochádza k neželaným prekvapeniam. Aj pre tieto dôvody je súčasným trendom modelovanie a analýza v projektovej fáze. Správnou analýzou je možné odhaliť mnohé nedostatky, ktoré sú v projektovej fáze jednoducho odstrániteľné. Analýzou dosiahnuteľnosti stavov výrobných liniek je možné overiť požadované vlastnosti a flexibilitu výrobnej linky. V prípade, že želané stavy nie sú dosiahnuteľné je potrebné urobiť zmeny v návrhu. Pre dosiahnuteľné stavy je možné vykonať ďalšie analýzy času výroby, výkonnosti alebo nákladov na výrobu.

POĎAKOVANIE

Táto práca bola vytvorená realizáciou projektu Rozvoj Centra informačných a komunikačných technológií pre znalostné systémy (kód ITMS projektu: 26220120030) na základe podpory operačného programu Výskum a vývoj financovaného z Európskeho fondu regionálneho rozvoja.

Táto práca bola podporená z vedeckého projektu Vega č. 1/0286/11 Grantovej agentúry SR pod názvom „Dynamické hybridné architektúry v multiagentových sieťových riadiacich systémoch“.

LITERATÚRA

[1] S. Bajzátová, Rôzne pohľady na Moore-Penrosovu pseudoinverziu: Diplomová práca. Brno: Masarykova univerzita, 2006. 57 s.
[2] M. Čopík, Návrh a realizácia riadenia pružného výrobného systému: Diplomová práca. Košice: TU, 2010. 75 s.
[3] M. Čopík, S. Laciňák, J. Jadlovský, Návrh a realizácia riadenia pružnej výrobnej linky. In Electrical Engineering and Informatics: Proceeding of the Faculty of Electrical Engineering and Informatics of the Technical University of Košice [CD-ROM]. Košice : Technická univerzita, 2010 s. 561-564. ISBN 978-80-553-0460-1.
[4] Š. Hudák, Decompositional Reachability Analysis. Journal of ElectricalEngineering, Vol. 45, No. 11, 1994 pp.424-431.
[5] Š. Hudák, New Approach to Solving the Reachability Problem: on the way to methodology of the analysis of time-critical systems. DrSc thesis, Národná univerzita T. Ševčenka, Kijev, Ukrajina, 1995, 270 pp.
[6] J. Ligušová, Metódy riadenia eliminujúce vplyv oneskorení spôsobených komunikačnou sieťou v sieťových riadiacich systémoch: Dizertačná práca. Košice: TU, 2003. 143 s
[7] V. Kordic, Petri net theory and applications, Vienna: I-Tech Education and Publishing, 2008. 544 s. ISBN 978-3-902613-12-7

Detekcia tvárí s využitím neurónových sietí

Dušan DOVALA, Rudolf JAKŠA

Katedra kybernetiky a umelej inteligencie, Fakulta elektrotechniky a informatiky,
Technická univerzita v Košiciach, Slovenská republika

dusandovala@gmail.com, jaksa@neuron.tuke.sk

***Abstrakt* — Táto práca sa zaoberá skúmaním využitia Kohonenovej neurónovej siete pri detekcií ľudských tvárí v obraze. V úvode práca popisuje problematiku detekcie tvárí a rôzne metódy detekcie tvárí využívajúcich neurónové siete. V hlavnej časti sa zaoberá návrhom a implementáciou systému pre detekciu tvárí pomocou Kohonenových sieti. Experimentálne overuje možnosti použitia navrhnutej metódy.**

***Kľúčové slová* — detekcia tvárí, neurónová sieť**

I. ÚVOD

V práci vychádzame z teórie o Kohonenových sieťach. Použili sme Kohonenovú neurónovú sieť kvôli jej schopnosti zhlukovej analýzy. Ďalším dôvodom pre použitie tohto typu neurónovej siete bol môj záujem o metódy nekontrolovaného učenia. Systém je vytvorený v programovacom jazyku C# s využitím frameworku .NET vo vývojovom prostredí Microsoft Visual Studion 2010. Celkovo sme pri testovaní systému vykonali šesť experimentov, počas ktorých sme sa snažili analyzovať schopnosť systému detekovať tvár v obraze. Pri návrhu a testovaní systému sme postupovali tak, že sme vytvorili základný návrh systému, ktorý sme experimentálne otestovali, a na základe výsledkov z týchto testov sme sa snažili použitím rôznych techník a metód upraviť systém tak, aby sme dosiahli zlepšenia výsledkov detekcie tváre. Pri experimentoch sme použili rôzne testovacie obrázky, aby sme otestovali správanie sa systému pri rôznej komplexnosti pozadia, ako aj chovanie pri rôznych svetelných podmienkach scény.

Detekcia tvári je prvá etapa automatických systémov pre rozpoznávanie tvárí, pretože tvár musí byť lokalizovaná vo vstupnom obraze pred tým, ako sa ju snažíme zaklasifikovať. Detekcia tvárí je zvyčajné dvojfázový proces: najskôr je preskúmaný celý obrázok za účelom identifikovať oblasti ktoré sú označené ako tvár. V druhej fáze sa získava presnejší odhad umiestnenia a veľkosť tváre[1].

Algoritmy pre detekciu tvárí môžu byť rozdelené do štyroch kategórií:

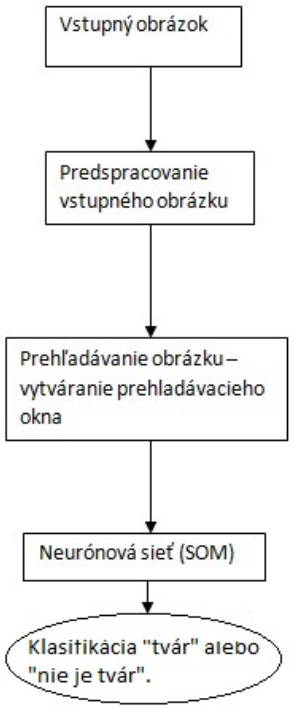
1. Knowledge-based metódy, ktoré sú založené na poznatkoch o geometrií a usporiadaný rysov ľudských tvárí.
2. Prístup nemenných znakov sa usiluje o nájdenie takých štrukturálnych vlastností, ktoré existujú, aj keď sa menia napr. svetelné podmienky.
3. Template-matching, je to metóda ktorá na detekciu tvárí využíva rôzne filtre a masky pomocou ktorých sa snaží nájsť tvár v obraze.
4. Appearence-based metódy. Kým metódy template-matching sa spoliehajú na preddefinované šablóny alebo modely, appearence- based metódy používajú veľké množstvo príkladov, ktoré zahŕňajú rôzne varianty (farbu očí, farbu pleti, tvar ....).

II. NÁVRH A IMPLEMENTÁCIA SYSTÉMU PRE DETEKCIU TVÁRÍ S VYUŽITÍM NEURÓNOVÝ SIETÍ

V tejto práci sa venujeme návrhu systému na detekciu tvárí v obraze pomocou neurónových sieti a to konkrétne pomocou samoorganizujúcej sa mapy (SOM) tiež známej ako Kohonenová neurónová sieť.

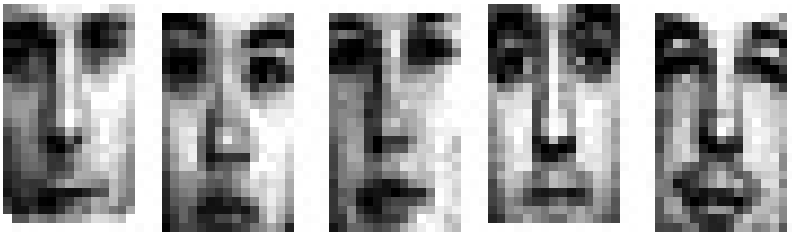
Prvým krokom pri detekcií tvárí je predspracovanie vstupného obrázku. Samotné predspracovanie spočíva v troch krokoch.

1. Zmena rozlíšenia na 640x480
2. Transformácia z RGB formátu do grayscale (čiernobiely) formátu.
3. Ekvalizácia histogramu.

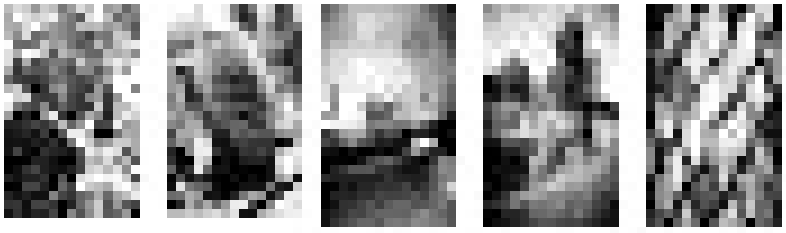


Obr. 1 Základná schéma systému. Predspracovanie: Zmena vstupného obrázku na štandardné rozlíšenie 640x480 transformácia obrázku z RGB formátu do grayscale (čiernobiely) formátu a ekvalizácia histogramu. Prehľadávanie obrázku: pre detekciu tvárí vo vstupnom obrázku musíme tento obrázok rozdeliť na menšie časti (oblasti) tzv. kandidátov ktoré reprezentujú určité oblasti obrázku. Týmto rozdelením vlastne definujeme okno prehľadávanie ktorým segmentujeme vstupný obrázok na menšie oblasti. Neurónová sieť: samotná neurónová sieť zabezpečuje klasifikáciu oblasti do dvoch tried a to "tvár" alebo "nie je tvár".

Samotná neurónová sieť pozostáva z tridsaťšesť neurónov, kde každý jeden neurón predstavuje jednu z klasifikačných tried. Pri učení neurónovej siete sme použili trénovaciu množinu, ktorá pozostáva zo 146 obrázkov, ktoré sú rovnomerne rozdelené na obrázky reprezentujúce ľudskú tvár, a na obrázky ktoré nereprezentujú tvár. Teda máme 73 obrázkov tvári a 73 obrázkov bez tvári. Databáza vychádza z databázy, ktorá bola použitá v práci [3]. Pre potreby tejto práce bola upravená a to tak, že boli odstránené oblasti ktoré nezodpovedali tvárí. Celá databáza bola upravená pomocou ekvalizácie histogramu, vďaka čomu došlo k vyrovnaniu jasových hodnôt jednotlivých obrázkov. Ďalej došlo k úprave rozlíšenia na rozmery 15x25. Veľkosť rozlíšenia sme zvolili tak, aby obrázky neboli zbytočne veľké, čo by malo za následok zvýšenie výpočtovej náročnosti, ale zároveň aby neboli príliš malé, čo by mohlo spôsobiť stratu tvárových rysov, na základe ktorých bude systém ďalej detekovať tváre.

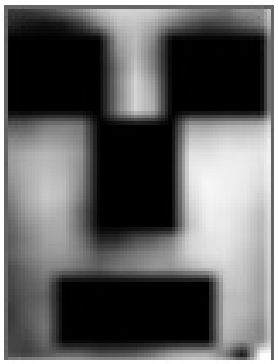


Obr. 2 Zobrazenie časti databázy po úprave. V databáze sa celkovo vyskytuje 73 obrázkov reprezentujúcich tvár z celkového počtu 146 obrázkov.

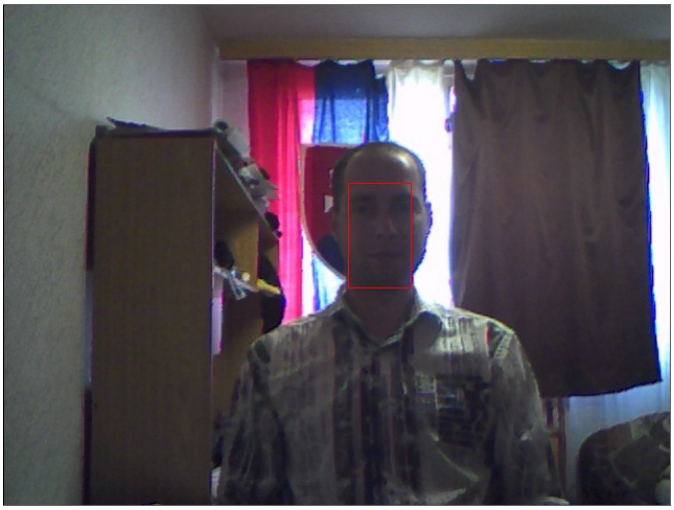


Obr. 3 Ukážka trénovacej množiny pre obrázky nereprezentujúce tvár. Obrázky boli poskladané z viacerých typov, prevažne sme uprednostňovali textúry. V databáze sa vyskytuje 73 obrázkov ktoré nereprezentujú tvár z celkového počtu 146 obrázkov.

Pri experimentoch sme použili viacero modifikácií systému za účelom zlepšenia schopnosti systému detekovať tvár. Najlepšie výsledky boli dosiahnuté pomocou zredukovania príznakového priestoru (viď Obr.4) a pomocou prahovania neurónu. Pri tejto metóde pri samotnej klasifikácií nesledujeme len to ktorý neurón zareagoval na daný vstup neurónovej siete, ale aj to do akej miery. Inak povedané neuvažujeme automaticky že sa jedná o tvár ale iba o oblasť v ktorej sa pravdepodobne tvár nachádza. O tom či daná oblasť je alebo nie je tvár, rozhodujeme na základe vopred definovaného prahu neurónu.



Obr. 4 Znázornenie špecifikovaných oblasti príznakového priestoru. Čiernou farbou sú znázornené pixely ktoré predstavujú oblasti ktoré reprezentujú príznakový priestor. Jednotlivé pixely boli vyberané na základe snahy vytvoriť príznakový priestor ktorý obsahuje rysy tváre ako sú oči, ústa a nos.



Obr. 4 Výstup zo systému po prahovaní neurónu, hodnota neurónu musí byť z intervalu <min, max> kde min predstavuje minimálnu hodnotu prahu a max naopak maximálnu hodnotu prahu neurónu.

Systém sa javí ako veľmi citlivý na presne umiestnenie tváre do okná prehľadávania. Pokiaľ dôjde iba k čiastočnému zachyteniu tváre oknom prehľadávania, neurónová sieť nie je schopná klasifikovať daný vstup ako tvár. Tento nedostatok môže byť spôsobený trénovacou množinou kde sa nachádzajú iba príklady tvári umiestnené viac menej v strede obrázka. Teda neurónová sieť neobsahuje neurón, ktorý by reprezentoval iba čiastočné zachytenie tváre či už na x-ovej alebo y-ovej osi.

Pri poslednom experimente sme sa rozhodli použiť obrázok na ktorom sa vyskytuje mnoho tvári. Zvolili sme taký obrázok, aby boli niektoré tváre zobrazené na obrázku pod rôznymi uhlami zhruba v rovnakej vzdialenosti od snímacieho zariadenia.



Obr. 5 Výsledok experimentu č.6, Výsledkom tohto experimentu bolo nájdenie jednej tváre z pätnástich ktoré boli zobrazené na vstupnom obrázku.

Výsledkom tohto experimentu bolo nájdenie jednej tváre z pätnástich ktoré boli zobrazené na vstupnom obrázku. Neúspech detekcie všetkých tvárí mohol byť spôsobený viacerými faktormi, ako je natočenie tváre pod určitým uhlom voči snímaciemu zariadeniu, citlivosť neurónovej siete na presné umiestnenie tváre v okne prehľadávania , problém nastavovania prahu neurónu, a v neposlednom rade sa ukázalo, že systém je do značnej miery závislý od svetelných podmienok scény, aj keď v práci využívame princíp ekvalizácie histogramu za účelom rovnomerného rozloženia jasových hodnôt, tak tento prístup nezabezpečí rovnaké podmienky pre rôzne testovacie obrázky.

III. ZÁVER

Na základe vykonaných experimentov s Kohonenovou neurónovou sieťou pre detekciu tvárí na experimentálnej dátovej množine môžeme skonštatovať, že systém má potenciál na detekovanie tvárí, avšak úspešnosť nie je dostačujúca na praktické využitie. Implementácia navrhnutej metódy detekcie tvárí nie je náročná. Jednotlivé úpravy metódy, ktoré sme vykonali počas experimentov priniesli zlepšenie vlastností systému ale nie natoľko aby bola navrhnutá metóda použiteľná v praxi. Počas vykonávania experimentov sme zistili zopár nedostatkov ktoré spôsobujú nízku úspešnosť korektnej detekcie.

POĎAKOVANIE

Táto práca bola vytvorená realizáciou projektu Rozvoj Centra informačných a komunikačných technológií pre znalostné systémy (kód ITMS projektu: 26220120030) na základe podpory operačného programu Výskum a vývoj financovaného z Európskeho fondu regionálneho rozvoja.

LITERATÚRA

[1] Abdallah S. Abdallah, A. Lynn Abott, Mohamad Abou El-Nasr: A New Face Detection Technique using 2D DCT and Self Organizing feature Map. World Academy of Science, Engineering and Technology 27 2007.

[2] Beszédeš, Marian: Detekcia tvárí pomocou neurónovej siete. Bratislava: Slovenská technická univerzita, Fakulta elektrotechniky a informatiky.

[3] Dovala Dušan: Použitie Kohonenovej siete pri rozpoznávaní tvárí. Košice: Technická univerzita v Košiciach, Fakulta elektrotechniky a informatiky, 2009. 49s.

[4] Filareti Tsalakanidou, Sotiris Malassiotis 2 and Michael G. Strintzis:FaceRecognition-Face Detection,Global Approaches for, Feature Based Techniques, Problems and Considerations, Conclusions and Feature Developments. Aristotle University of Thessaloniki, Thessaloniki, Greece. Informatics and Telematics Institute, Centre for Research and Technology Hellas, Thessaloniki, Greece.

[5] Inseog, Kim, Joon Hyung Shim, Jinkyu Yang: Face Detection.

[6] Kalavdekar Prakash N.: Face Detection using Neural Network. International journal of Computer Applications (0975 - 8887), volume1 - No. 14.

[7] Sinčák Peter – Anderjková Gabriela: Neurónové siete Inžiniersky prístup (1. diel): 22 Júla 1996.

Towards the optimization of IT service delivery processes in governmental environment

Jana DUCHOVÁ, Karol FURDÍK

Department of Cybernetics and Artificial Intelligence, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Slovak Republic

janka.duchova@gmail.com, karol.furdik@tuke.sk

***Abstract* — The paper presents a contribution to the knowledge of business process analysis, standardization, modeling and design solutions applied to optimize IT services. The process of submitting of electronic tax declarations was selected as a subject of the analysis. The submission of tax declarations is an IT service provided by the financial government by means of the Tax Directorate of the Slovak Republic. The focus is given on the processes of service delivery according to ITIL v3, which were used for an extension and optimization of processes that are employed on the service provider side. The goal was to eliminate the service failures and improve the communication with users. The proposal is demonstrated on formal models developed for existing and updated processes of the service.**

***Keywords* — process, standard, norm, government**

I. INTRODUCTION

The shift towards the process-oriented and service-based solutions is a trend that can be seen in the applied information technology (IT) in the last decade. The IT infrastructure providing services integrated into complex processes enables a close relationship and tight contract-based cooperation between service providers and customers by bringing an added value for both the actors. The paradigm of SAAS, i.e., the "software as a service", supported in enterprise IT applications by business process management and proper IT governance, is referenced in [1] as fourth wave of IT, which extends the mainframe systems, separate IT departments, and networked IT infrastructure of previous waves. The concept of services and underlying business processes enables creation and maintenance of flexible, distributed, and highly scalable solutions; however, the qualitative advancement of this approach is the possibility to measure and continuously monitor the performance and business value of IT services from both the provider and customer perspective.

The SAAS paradigm may invoke a change in the strategic goals of any larger organization (e.g. a business enterprise, governmental institution, or non-profit organization) towards the transformation of an organizational unit to the center of supporting services [2]. The services in such a center, including IT services, are established and maintained to support specified business goals and primary organizational processes. The necessary prerequisite is that the processes are well defined, established and, in an ideal case, properly documented. It implies that the *Business Process Management* (BPM), supported by some of process modeling techniques [3], is crucial for the control and maintenance of underlying processes in an environment of applied IT services [4]. Furthermore, the issues of strategy, change handling, application and process integration, as well as IT service operation need to be planned, managed, and aligned to organizational business goals, taking into account the technology constraints of an existing IT architecture. All these aspects are covered by the *IT Service Management* (ITSM), which is dealing with the design, deployment, and continual support of IT services during the whole service life cycle, in accordance with business requirements of an organization [2]. Consequently, ITSM together with model-based BPM can be employed as a powerful framework for optimizing the quality and effectiveness of provided (or even consumed) IT services in an environment built on the SAAS paradigm.

The content in the next sections is organized as follows. A brief overview of basic ITSM principles, together with the outline of IT service delivery processes is presented in section 2. The status and level of adopting ITSM in e-Government solutions in Slovakia is described in section 3. Ways of optimizing the maintenance of e-Government applications according to the ITSM processes of service delivery are discussed in section 4. In particular, the approach is demonstrated on an example of the Tax Directorate portal for on-line submission of tax documents. Section 5 concludes the topic with a summary and some hints towards possible future advances of ITSM in e-Government.

II. ITIL AND SERVICE DELIVERY PROCESSES

The ITSM is nowadays supported by several frameworks that provide a methodology, specification of required processes, and implementation guidelines for managing IT services and infrastructure on several hierarchical levels [2]. Frameworks such as COBIT (www.isaca.org), MSP (www.msp-officialsite.com), or PRINCE2 (www.prince2.com) are focused on upper levels of IT Governance, strategic IT management, and project-based management of changes. On the operational level of IT services, the IT Infrastructure Library (ITIL, www.itil-officialsite.com) and the related standard ISO/IEC 20000 are world leading and nowadays the most frequently applied frameworks. ITIL is a collection of best practices, composed of five books of Service Strategy, Service Design, Service Transition, Service Operation, and Continual Service Improvement, which correspond to the respective phases of IT service life cycle [5]. Currently, ITIL is available in version 3, released in year 2007 and updated in 2011. The ISO/IEC 20000 standard is based on previous version of ITIL v2 [6] and provides a code of practice, evaluation metrics, and certification schema for ITSM on its operational level.

ITIL and ISO/IEC 20000 specify the processes of IT service delivery, which are essential for optimizing the service operation maintenance, achieving required quality of the service, and increasing the service provision efficiency from the perspective of both the providers and customers. The IT service delivery processes are specified in ITIL v3 [5] as follows:

- *Service Level Management* (SLM) is the process of negotiating Service Level Agreements (SLA), and ensuring that these are met. SLM is responsible for ensuring that all IT service management processes, internal Operational Level Agreements (OLA) and contracts with external suppliers are appropriate for the agreed service level targets. SLM monitors and reports on service levels, and holds regular customer reviews.
- *Availability Management* (AM) is the process responsible for defining, analyzing, planning, measuring and improving all aspects of the IT service availability. It is responsible for ensuring that all IT infrastructure components, processes, tools, roles, etc., are in line with the agreed service level targets for availability.
- *Capacity Management* (CM) is responsible for ensuring that the capacity of IT services and the related IT infrastructure is able to deliver agreed service level targets in a cost effective and timely manner. CM considers all resources required to deliver the IT service as whole, and plans the infrastructure updates for short, medium, and long term business requirements.
- *IT Service Continuity Management* (ITSCM) is handles the management of risks that could seriously impact IT services. ITSCM ensures that the IT service provider can always provide minimum agreed service levels, by reducing the risk to an acceptable level and planning for the recovery of IT services. ITSCM should be designed to support the Business Continuity Management.
- *IT Financial Management* (ITFM) covers functions and processes responsible for managing the budgeting, accounting and charging requirements of IT service providers.

Furthermore, the *Service Desk* is required as a single point of contact between the service provider and the users. A typical Service Desk manages incidents and service requests, and also handles all the communication with the IT service users.

III. IT SERVICE MANAGEMENT APPLIED IN E-GOVERNMENT IN SLOVAKIA

ITSM, and ITIL in particular, has been applied in numerous mostly business enterprises in past few years, with a tendency of rapid growth. In Slovakia, this trend is documented, for example, in [7]. In 2005, the concept of ITIL was known for representatives of only 12,1% of Slovak enterprises and organizations; however, the interest on ITSM and ITIL methods and the intention to implement some of the ITSM processes in near future was indicated was relatively high (over 50%). At the beginning of 2010, the familiarity with ITSM/ITIL increased

significantly and was over 73% [8]. This rate is comparable with countries as Germany or Austria; in Scandinavian countries the rate is on about 10% higher.

The survey presented in [8] provides a list of the most frequently implemented ITSM/ITIL processes (depicted in Figure 1), which include namely the end-user functions and processes such as Service Desk and Incident Management, followed by internal processes of Change Management and Service Asset and Configuration Management. The processes of Access Management and Request Fulfillment became popular for the automated processing of user requests. The study also presents the list of processes that organizations plan to implement in near future. It includes the management of the service catalog, service portfolio, availability, capacity, demands, finances, suppliers, etc.

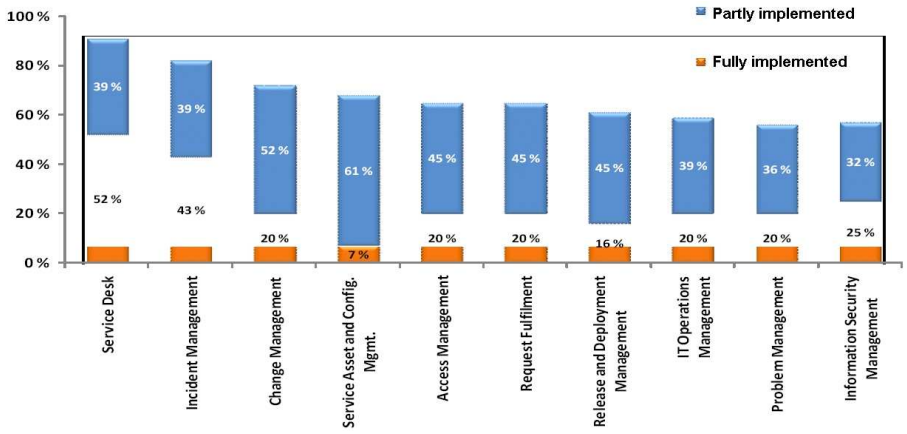


Fig. 1 Chart of fully or partly implemented ITIL processes in year 2010 [8].

The adoption of ITSM principles and implementation of ITIL processes is ongoing increasingly in business-oriented organizations such as banks and other financial institutions, insurance companies, or larger IT solutions providers. However, the advantages of ITSM were recognized in non-profit and governmental institutions as well, although the real usage of ITSM/ITIL in practice is not so advanced.

Table 1
Proposed time frame of implementing ITIL in e-Government IT services in Slovakia [9]

Activity	Date of completing the implementation
State-of-the-Art analysis	4Q[a] / 2006
Proposal of the goal state and of the implementation steps	2Q / 2007
Configuration Management	4Q / 2007
Service Desk, Incident and Problem Management	2Q / 2008
Change Management, Release Management, deployment of the Continual Service Improvement	3Q / 2008
Service Level Management, Availability Management	4Q / 2008
Continuity Management, Capacity Management, IT Financial Management	3Q / 2009

[a] Q = quarter of the year.

The IT infrastructure is mostly employed in municipalities and governmental institutions for e-Government solutions that enable electronic communication and information exchange between citizens and government, or internally between governmental institutions [10]. Assuming the SAAS paradigm of maintaining and provisioning the IT-enabled e-Government solutions, the respective applications should be managed as IT services, in accordance with ITSM/ITIL best practices and recommendations. In Slovakia, ITIL was introduced as a promising ITSM approach in year 2007, on the 6th annual conference ITAPA (www.itapa.sk). Dr. Hlavačka, representing the Tax Directorate of the Slovak Republic (www.drsr.sk), stated in his presentation [9] that "*ITIL is both a necessity and competitive advantage in business area*" and thus "*the need of applying ITSM is increasing with respect to the grow of e-Government*

services and with the increasing dependency of organizations on an IT support". The time frame of implementing ITIL in years 2006-2009, presented in Table 1, was proposed in [9] as well. We have contacted Dr. Hlavačka via e-mail; however, we were not able to find out how and to which extent the proposed time frame was fulfilled.

In year 2011, upon a supervision of the Ministry of Finance of the Slovak Republic, the document [11] was elaborated by the KPMG Slovensko spol. s.r.o. advisory company and issued as a basic methodology framework for the informatization of governmental and public institutions. The framework applies the ITIL recommendations for the management and operation of the Governmental Information System (Informačný systém verejnej správy, ISVS). The methodology defines rules for settings and documenting the processes of IT organizations that provide particular solutions in e-Government. Based on ITIL v2 recommendations, the document introduces basic contracts on the provision of IT services with agreed qualitative and quantitative levels; in other words, the methodology specifies the SLA contracts between the Ministry of Finance and the maintainer of ISVS.

IV. Use Case of Submitting Electronic Documents

To investigate the possibilities for optimizing an e-Government IT service in accordance with ITIL best practices, we have selected the service of on-line submission of tax declarations and other types of electronic documents. The service is provided by the Tax Directorate of the Slovak Republic and is accessible at www.drsr.sk.

Since we had no access to the internal data and information on processes, IT infrastructure, and other settings of the service, we have analyzed the web portal from the citizen / customer perspective. Furthermore, we have contacted the Tax Directorate Headquarters in Banská Bystrica with the request of providing the information, based on the regulation No. 3/2010/250. We have obtained the User Manual for authorized electronic services [12] and the list of frequently asked questions that users addressed to the application providers [13]. In addition, we have regularly followed and investigated the situation related to the on-line submission of tax declarations, together with the feedback provided by various users on the Internet. Based on all these materials, we have accomplished a detailed analysis of the current status of the service, which resulted in an implementation of formal models for all identified processes of the service delivery. Top-level model of the overall process of submitting electronic documents is accompanied with partial models of registration, authorization, user accounts management, submission of documents with and without electronic signature, and electronic filling room. A fragment of the top-level process model is presented in Figure 2. The whole set of BPMN process models (www.bpmn.org), developed using the QPR ProcesGuide tool, is included as the Appendix D in [14].

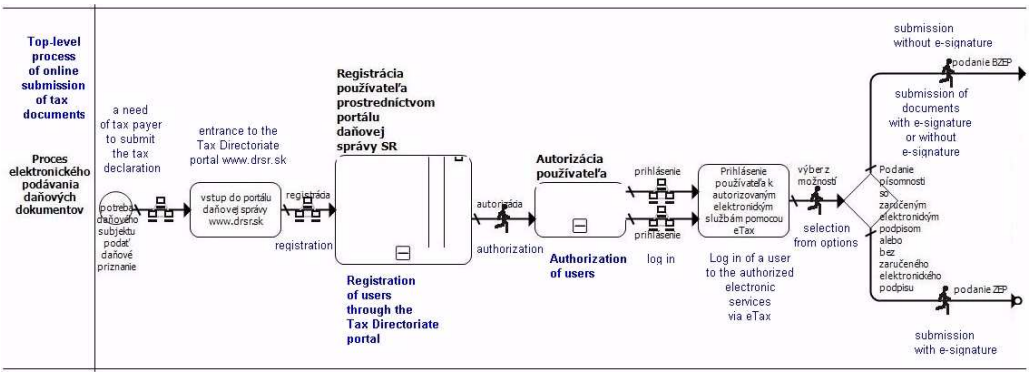


Fig. 2 The first part of the top-level process model of submitting tax documents, the "as is" status.

Analysis of user inputs, questions, and requirements that tax payers addressed to the service provider resulted in a set of identified issues, which were taken as the main indicators for the process optimization and service improvement. The issues were prioritized and grouped into three sub-sets according to the importance and influence on the overall process. Two issues with the highest priority were identified as follows:

- Notice for authorized users: The *User Accounts Management* sub-service is not accessible. Reason of the failure: Increasing number of on-line users can cause problems with functionality of some of the portal modules and, if combined with other factors, may lead to the unavailability of the whole portal.

- Notice for authorized users: The Tax Directorate, together with relevant suppliers, works intensively on the provision of technical conditions for stable service operation of the www.drsr.sk portal. Till the implementation of necessary updates, delayed responses or temporary breakdowns may occur. Reason of the failure: Planned maintenance of the portal is scheduled for some days between 10:00 PM and 11:00 PM. (However, the schedule of planned maintenance for year 2012 was not available on the portal even in March 2012.)

The issues with lower priority were taken as supportive indicators for specifying more precisely the failure reasons of the two highest priority issues.

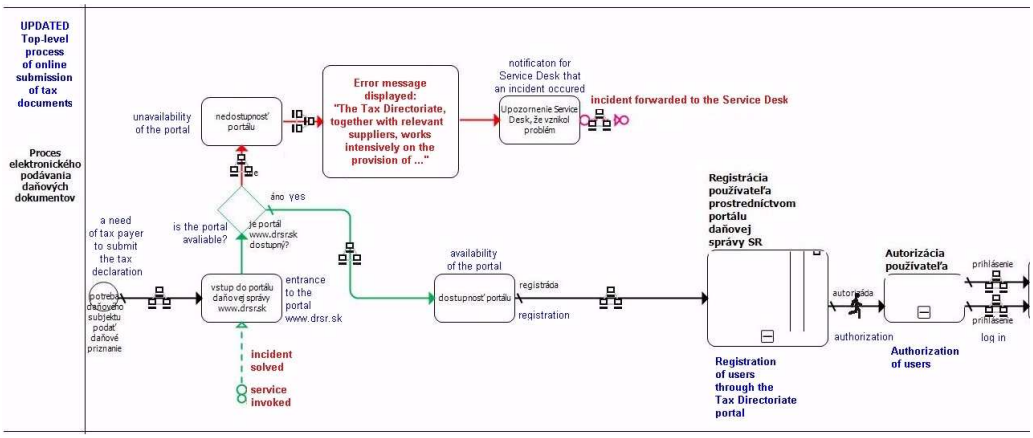


Fig. 3 The optimized top-level process model of submitting tax documents, the "to be" status.

Optimizing solution for all the identified issues was proposed in accordance with ITIL best practices. Namely, the IT service delivery processes, which were outlined in section 2 above, were adapted for the process of tax documents submission in the Tax Directorate portal. The supportive ITIL processes of Configuration, Incident, Problem, Change, and Release managements were specified and included into the BPMN process model of the service. For example, the optimized process model that eliminates the second top-priority issue is presented in Figure 3. The Service Desk sub-process, consisting of three escalation levels, was included to handle the management of problems with the service availability during planned or occasional system maintenance, as well as various other types of incidents or change requests. The first issue of accessibility failure caused by the increasing number of on-line users was identified as rather more complex one; the optimizing solution includes all the service delivery processes that are recommended by ITIL. The respective process model is presented in the Appendix F of [14].

V. Conclusion

The Tax Directorate portal enabling on-line submission of tax documents was not accessible during a peak season this year, i.e. in March 25-27, 2012 [15]. G. Dianová, the speaker of the Tax Directorate institution, stated that "*the existing portal www.drsr.sk is not able to handle the pressure before the tax submission deadlines, because the strategic as well as technological development of the portal was heavily ignored in years 2006-2010*". We are quite convinced that these deficiencies are related to the capacity and availability of the portal as an IT service, which can be solved by proper implementation of ITSM methods provided, for example, by ITIL or similar framework. In this paper we have tried to present an approach of updating and optimizing processes that drive the IT service maintenance by a consistent application of processes proposed in ITIL for the service delivery. The presented proposal, even limited by lack of knowledge of underlying technical details of the portal, can serve as a starting point for improvement and continuous sustainable maintenance of the Tax Directorate portal as an IT service that is capable to bring added value for both the provider and users.

Acknowledgment

The work presented in this paper was supported by the Slovak Cultural and Educational Grant Agency of the Ministry of Education of the Slovak Republic under the contract No. 065TUKE-4/2011.

REFERENCES

[1] R. Lenčeš, "Ako zlepšiť governance IT služieb," *eFocus*, vol. VII, 3/2008, pp. 22–25.
[2] R. Janáč, "Význam BPM pre riadenie IT," *eFocus*, vol. VII, 3/2008, pp. 30–34.
[3] L. Révészová, D. Paľová, "Základy modelovania podnikových procesov," Technická univerzita v Košiciach, Ekonomická fakulta, 2009.
[4] J. Sabol, "Predmet: Informačné systémy podnikových procesov," Použité prednášky: Procesný prístup a procesné riadenie, Implementácia procesného riadenia, Princípy procesného modelovania, Technická univerzita v Košiciach, Fakulta elektroniky a informatiky, unpublished.
[5] Office of Government Commerce OGC, "ITIL - Service Design," TSO, The Stationery Office, UK, 2007.
[6] M. Havelka, "Hlavné rozdiely ITIL V3 verzus ITIL V2," *eFocus*, vol. VII, 3/2008, pp. 12–14.
[7] ITSM, "IT Service Management, ITIL®V3, ISO/IEC 20000," [online] [cit. 10.5.2012], available at: www.itsm.sk.
[8] R. Bělina, "Vítězné tažení IT Service Managementu postaveného na ITIL v České republice a na Slovensku," *eFocus*, vol. IX, 2/2010, pp. 21–24.
[9] J. Hlavačka, "Implementácia ITIL vo verejnej správe," ITAPA 2007, Bratislava, 13.11. 2007 [online] [cit. 10.5.2012], available at: http://www.itapa.sk/data/att/1791_subor_sk.pdf.
[10] O. Škvarčeková, "E-Government na Slovensku," *Nová ekonomika*, vol. 3, 3/2004, pp. 242–245.
[11] KPMG Slovensko spol.s.r.o., "Riadenie prevádzky informačných technológií," Ministerstvo financií Slovenskej republiky, [online] [cit. 14.5.2012], available at: www.informatizacia.sk.
[12] Daňové riaditeľstvo Slovenskej republiky, "Používateľská príručka pre Autorizované elektronické služby," [online] [cit. 12.5.2012], available at: https://www.drsr.sk/drsr/ur/doc/pouziv_prirucka.pdf.
[13] Daňové riaditeľstvo Slovenskej republiky, "Často kladené otázky a problémy v súvislosti s využívaním elektronických služieb," [online] [cit. 14.5.2012], available at: www.drsr.sk.
[14] J. Duchová, "Modelovanie a optimalizácia podnikových procesov vo fáze dodávania IT služby," Diplomová práca, Technická univerzita v Košiciach, Fakulta elektroniky a informatiky, 2012.
[15] eTrend, "Portál daňovej správy bol niekoľko dní nedostupný, Priznania by sa mali dať posielať internetom aj počas výpadkov," TREND Holding s.r.o., published 28.3.2011, [online] [cit. 10.5.2012], available at: http://financie.etrend.sk/dane-a-odvody/portal-danovej-spravy-bol-niekolko-dni-nedostupny.html.

Riadenie robota pomocou platformy Android

Tomáš DUPLINSKÝ, Peter PAPCUN, Ján JADLOVSKÝ

Katedra kybernetiky a umelej inteligencie, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

tomas.duplinsky@student.tuke.sk, peter.papcun@tuke.sk, jan.jadlovsky@tuke.sk

***Abstrakt* — Predkladaný článok sa zaoberá riadením robota Mitshubishi RV – 2SDB využitím platformy Android. Článok sa zaoberá popisom komunikačných väzieb a vysvetleniu princípu riadenia robota v reálnom čase. Zároveň stručne charakterizujeme funkcionalitu navrhnutých aplikácií.**

***Kľúčové slová* — robot, riadenie, Android, bluetooth**

I. ÚVOD

Súčasnosť je charakterizovaná prudkým rozvojom informačných a telekomunikačných technológií. Zariadenia využívajúce operačný systém Android sú výkonné mobilné počítače s možnosťou stáleho pripojenia na internet a s bohatou paletou zabudovaných senzorov. Tieto zariadenia sú zaujímavé z hľadiska robotiky okrem iného aj pre množstvo komunikačných rozhraní, medzi ktoré patria WiFi, Bluetooth, GSM/UMTS a USB, a pre integrované senzory, ako akcelerometer, gyroskop, GPS. Na rozšírení inteligentných mobilných zariadení má svoj podiel jednak zdokonaľovanie a miniaturizácia technológií, a jednak rozvoj internetovej infraštruktúry. Android patrí medzi najpopulárnejšie mobilné operačné systémy predovšetkým pre jeho výkonnostné kapacity a otvorenú architektúru.

Neustály vývoj v oblasti komunikačných technológií a robotiky, vedie k otázkam možnosti vzájomného využitia v praxi. Predkladaný článok sa zameriava najmä na spojenie inteligentného telefónu s operačným systémom Android a využitie tohto spojenia na riadenie robota MELFA RV – 2SDB od spoločnosti Mitshubishi.

Cieľom článku je poukázať na vytvorenú aplikáciu pre operačný systém Android, ktorá umožňuje riadenie vzdialeného robota v reálnom čase a komunikačné väzby medzi prvkami systému. Článkom chceme poukázať na nové možnosti a prístupy k riadeniu robotických zariadení, ako aj na možnosti ponúkané robotom RV – 2SDB.

II. KOMUNIKAČNÉ VÄZBY

Popis systému

Systém, pre ktorý bola navrhnutá aplikácia pod OS Android s názvom RobotControl je znázornený na Obr. 1. Systém pozostáva z riadiacej jednotky CR1D – 700, ktorá je prepojená s robotom MELFA RV – 2SDB prostredníctvom špeciálneho kábla dodaného spoločnosťou Mitshubishi. Riadiaca jednotka komunikuje s lokálnym počítačom prostredníctvom rozhrania Ethernet konkrétne typu 1000Base - T s využitím protokolu UDP. Lokálny počítač zabezpečuje komunikáciu medzi riadiacou jednotkou robota a mobilným zariadením prostredníctvom aplikácie Jupiter a integrovaného modulu Ganymedes, pričom sa využíva Bluetooth komunikačné rozhranie s protokolom RFCOMM.

Riadenie robota v reálnom čase

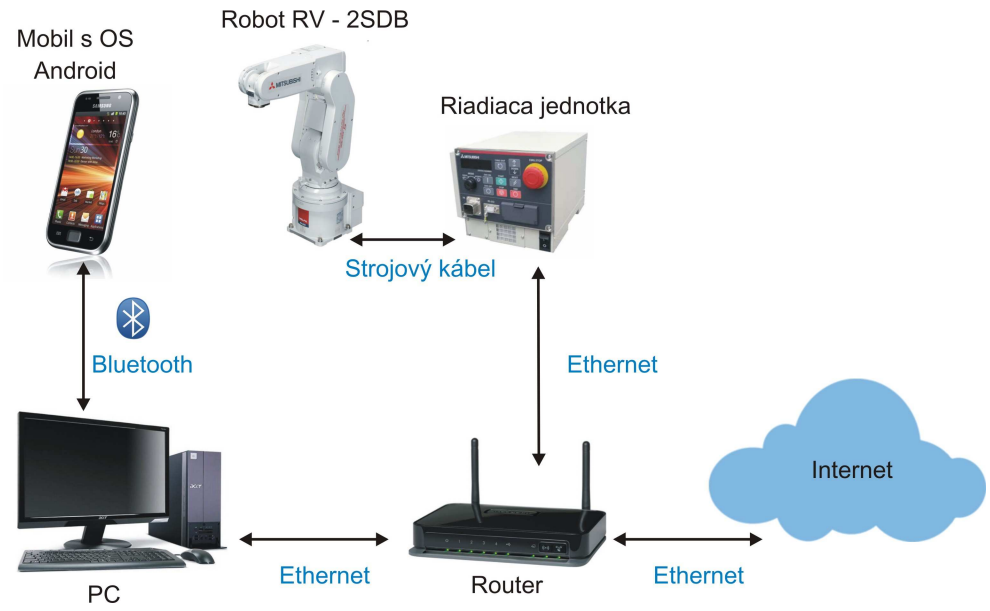
Programovací jazyk MELFA V pre riadiacu jednotku CR1D – 700 ponúka funkciou MXT (*Move External*). Táto funkcia umožnuje vzdialené riadenie robota v reálnom čase prostredníctvom komunikačného rozhranie Ethernet a protokolu UDP. Túto funkciu sme využili v rámci komunikácie medzi riadiacou jednotkou a lokálnym počítačom. V súvislosti s komunikáciou sme využili špeciálnu štruktúru komunikačného paketu, ktorá je definovaná spoločnosťou Mitshubishi.

Algoritmus komunikácie v reálnom čase

V prvom rade sa inicializuje komunikácia na strane riadiacej jednotky s príslušným programom obsahujúcim funkciu MXT. Pri spustení riadiaceho programu v riadiacej jednotke sa otvorí komunikačný port a vykoná sa inicializácia. Program následne čaká na spojenie a prichádzajúce dáta (štartovací paket) zo strany PC. Spustením vytvorenej aplikácie sa na strane PC inicializuje komunikácia využívajúca protokol UDP. Následne sa vytvorí štartovací paket, ktorý informuje riadiacu jednotku o prichádzajúcich riadiacich paketoch. Tento paket sa odošle do riadiacej jednotky.

Riadiaca jednotka príjme štartovací paket a odpovedá v podobe informácií o polohe robota, ktoré zasiela späť do PC. Typ kontrolných informácií o polohe robota, ktoré sa zasielajú späť do PC závisí od konfigurácie paketu. Prijatím potvrdzovacieho paketu (vrátane kontrolných dát z riadiacej jednotky) prebieha komunikácia rovnakým spôsobom s tým rozdielom, že sa mení typ paketu zo štartovacieho na riadiaci. To znamená, že po prijatí komunikačného paketu do riadiacej jednotky dochádza k vykonaniu pohybu ramena robota v závislosti od pozičných dát, ktoré sú uložené v prijatom riadiacom pakete. Následne riadiaca jednotka odpovedá aktuálnou polohou robotického ramena, v závislosti od konfigurácie komunikačného paketu a zvolených typov kontrolných dát, ktoré užívateľ nastaví podľa potrieb kontroly a monitorovania pozície robota.

Pred ukončením programu sa na strane PC vytvorí ukončovací paket, ktorý sa odošle do riadiacej jednotky, ktorú informuje o ukončení komunikácie. Táto komunikácia je jednosmerná, pretože riadiaca jednotka už neodpovedá na ukončovací paket. Následne sa ukončí program v riadiacej jednotke aj na strane PC.



Obr. 1 Schéma prepojenia jednotlivých prvkov systému

III. REALIZÁCIA ARCHITEKTÚRY SYSTÉMU

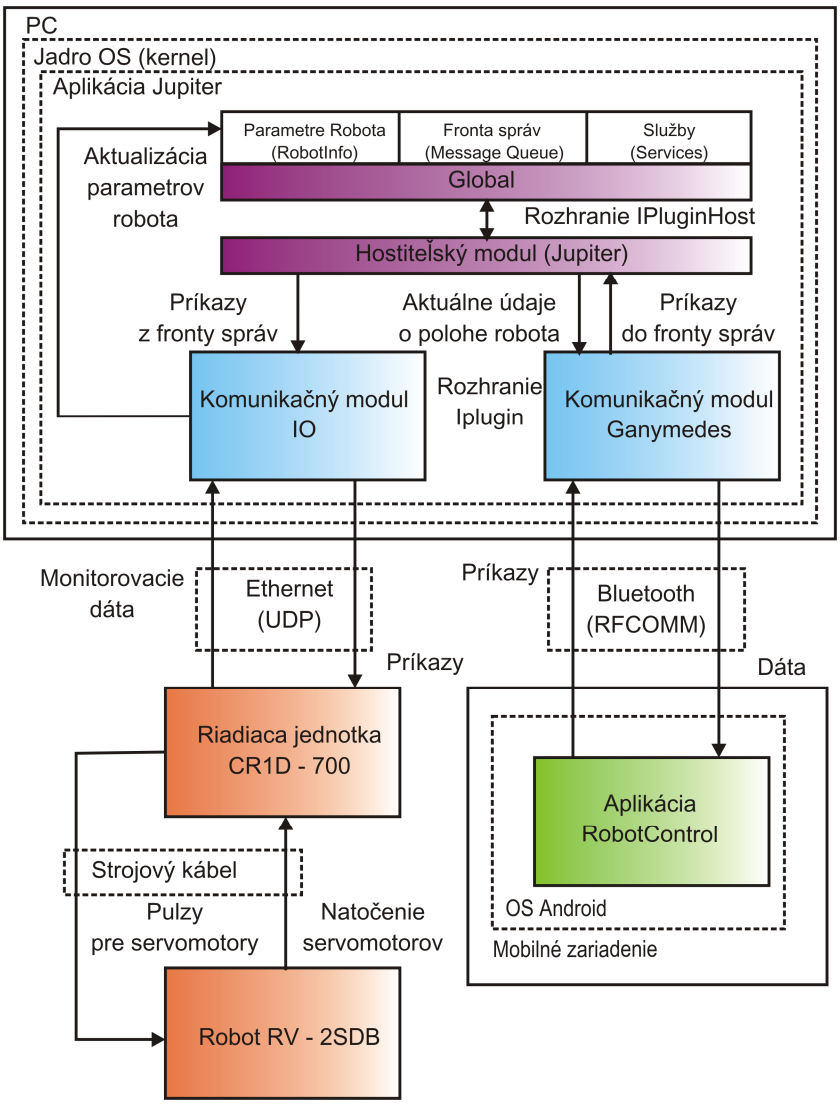
Pred samotnou realizáciou navrhnutých aplikácií bolo potrebné vyriešiť otázku, ako prebehne komunikácia a dátový tok medzi navrhnutými aplikáciami. Na Obr. 2 je znázornená bloková schéma komunikačného a dátového toku celého systému.

Z blokovej schémy (Obr. 2) je zrejme, že v systéme máme tri uzavreté obvody. Prvý uzavretý obvod predstavuje spätnoväzobná slučka medzi riadiacou jednotkou CR1DA – 700 a robotom RV – 2SDB. Riadiaca jednotka posiela pulzy pre jednotlivé servomotory robota pričom robot odpovedá riadiacej jednotke v podobe informácií o aktuálnom natočení jednotlivých servomotorov. Táto interná spätnoväzobná slučka je daná spoločnosťou Mitshubishi. Komunikácia medzi riadiacou jednotkou a robotom je realizovaná prostredníctvom strojového kábla, ktorý bol súčasťou dodávky od spoločnosti Mitshubishi.

Druhý uzavretý obvod pozostáva z komunikačného modulu Ganymedes a riadiacej jednotky robota. Komunikačný modul Ganymedes, vysiela riadiace povely na zmenu pozície robota prostredníctvom IO komunikačného modulu, pričom riadiaca jednotka informuje komunikačný modul o aktuálnej polohe robota. Komunikácia v tomto prípade prebieha prostredníctvom

Ethernetu.

Tretí uzavretý obvod pozostáva z komunikačného modulu Ganymedes a aplikácie RobotControl na strane mobilného zariadenia. Prostredníctvom aplikácie RobotControl užívateľ zadáva riadiace povely na zmenu pozície robota, ktoré sa odosielajú do komunikačného modulu Ganymedes. Následne tento modul poskytuje všetky potrebné dáta aplikácii RobotControl, ktoré sú nevyhnutné pre riadenie robota v reálnom čase. Komunikácia prebieha prostredníctvom bluetooth rozhrania.



Obr. 2 Komunikačný a dátový model

Popis funkčných blokov systému

Komunikačný systém Jupiter je navrhnutý pre zabezpečenie komunikácie PC s riadiacou jednotkou. Je to modulárny systém. Znamená to, že umožňuje definovať funkcie, ktoré následne môžeme „zabaliť“ do modulov, pričom tieto moduly fungujú ako nezávisle entity (pluginy) v systéme. Jednotlivé moduly komunikujú prostredníctvom hostiteľského modulu Jupiter, ktorý im poskytuje svoje služby (Služby, Fronta správ).

Komunikačný modul Ganymedes zabezpečuje komunikáciu a výmenu dát s mobilným zariadením. Modul zabezpečuje spracovanie všetkých príkazov a požiadaviek zo strany mobilného zariadenia a zároveň odpovedá na tieto požiadavky. Komunikačný modul umožňuje nahrávanie sekvencií pohybu robota do súborov. Tieto sekvencie je možné využiť pre iné aplikácie. Modul Ganymédes komunikuje s IO komunikačným modulom, ktorému predáva riadiace povely na zmenu pozície robota, ktoré prijíma zo strany mobilného zariadenia.

IO komunikačný modul zabezpečuje komunikáciu s riadiacou jednotkou, ktorá zahŕňa vytvorenie, udržiavanie a ukončenie komunikácie. Zabezpečuje odosielanie informačných alebo riadiacich paketov (odosiela štartovací, riadiaci a ukončovací paket). Modul vyčítavá príkazy z fronty správ (*Message Queue*) a posiela ich do riadiacej jednotky. Poskytuje možnosť

konfigurácie pripojenia. Aktualizuje globálnu triedu RobotInfo (parametre robota) na základe odpovedí z riadiacej jednotky.

Aplikácia RobotControl (Obr. 3) bola navrhnutá pre OS Android. Prostredníctvom tejto aplikácie je užívateľ schopný vzdialene riadiť robota v reálnom čase. Z pohľadu komunikácie môže aplikácia vystupovať v pozícií klienta ale aj servera. Aplikácia obsahuje grafické rozhranie pre definovanie typu riadenia (pozičné dáta, natočenie jednotlivých kĺbov robota). Pre samotné riadenie robota sú k dispozícií príslušné tlačidla, prostredníctvom ktorých môže užívateľ riadiť natočenie jednotlivých kĺbov robota. Informácie o aktuálnej polohe robota má užívateľ priamo na obrazovke mobilného zariadenia, vďaka čomu môže plnohodnotne riadiť pohyb robota. Aplikácia má integrovanú funkcionalitu pre vzdialené spúšťanie rôznych sekvencií pohybu robota.



Obr. 3 Aplikácia RobotControl

IV. ZÁVER

V prvom rade sa podarilo správne navrhnúť a realizovať komunikáciu medzi riadiacou jednotkou robota a mobilným zariadením. Využili sme komunikačný systém Jupiter a taktiež bluetooth komunikáciu s komunikačným protokolom RFCOMM. V rámci aplikácie Jupiter sa nám podarilo realizovať komunikačný subsystém Ganymedes, ktorý nám sprostredkúva komunikáciu medzi riadiacou jednotkou a mobilným zariadením.

Ďalším kľúčovým bodom bol návrh a realizácia aplikácie pod OS Android, ktorá poskytuje jednoduché a intuitívne rozhranie. Aplikácia ponúka plnohodnotné riadenie robota v reálnom čase. Poslednou charakteristickou črtou modulu Ganymeds je práca so sekvenciami pohybu robota. V komunikačnom module Ganymedes sa vytvorila funkcionalita, ktorá zabezpečuje nahrávanie sekvencie pohybu robota do súborov. Môžeme taktiež pracovať so súbormi a spúšťať uložené sekvencie zo strany modulu Ganymedes ako aj zo strany mobilného zariadenia. V súvislosti so spúšťaním sekvencií z mobilného zariadenia vidím uplatnenie takéhoto prístupu v oblastiach, kde je potrebné dynamický meniť výrobu a rýchlo meniť algoritmy v riadiacich jednotkách.

POĎAKOVANIE

Táto práca bola vytvorená realizáciou projektu Rozvoj Centra informačných a komunikačných technológií pre znalostné systémy (kód ITMS projektu: 26220120030) na základe podpory operačného programu Výskum a vývoj financovaného z Európskeho fondu regionálneho rozvoja.

LITERATÚRA

[1] DUPLINSKÝ, T.: Riadenie robota pomocou platformy Android, Diplomová práca, Košice: Technická univerzita v Košiciach, Fakulta elektrotechniky a informatiky, 2012, 94s.

[2] PAPCUN, P.: Riadenie robota integrovaného v pružnej výrobnej linke, Diplomová práca, Košice: Technická univerzita v Košiciach, Fakulta elektrotechniky a informatiky, 2011, 95s.

[3] ADAMONDY, P.: Riadenie robota rozpoznávaním ľudských pohybov pomocou zariadenia Kinect, Diplomová práca. Košice: Technická univerzita v Košiciach, Fakulta elektrotechniky a informatiky, 2012., 76s.

[4] PAPCUN, P. – ČOPÍK, M.: Remote control of Mitsubishi industrial robot, In: SCYR 2012: 12th Scientific Conference of Young Researchers of Faculty of Electrical Engineering and Informatics Technical University of Košice: FEI TU, 2012 - ISBN 978-80-553-0644-5.

Predikcia bankrotu firiem pomocou vhodných metód objavovania znalostí

Tomáš GROSS, František BABIČ

Katedra kybernetiky a umelej inteligencie, Fakulta elektrotechniky a informatiky,
Technická univerzita v Košiciach, Slovenská republika

tomas.gross@student.tuke.sk, frantisek.babic@tuke.sk

***Abstrakt* — Analýza finančnej situácie firiem predstavuje dôležitý zdroj informácii pre manažérov, investorov, dodávateľov alebo výkonných pracovníkov. V tomto prípade sme sa rozhodli predpovedať stav firmy (bankrotujúca alebo aktívna) na základe vstupných finančných ukazovateľov extrahovaných z finančnej databázy Amadeus. Pre tieto účely boli vybrané tri algoritmy dolovania dát, celý proces objavovania znalostí bol realizovaný na základe metodológie CRISP-DM. Praktické experimenty boli vykonané pomocou voľne dostupného nástroja RapidMiner. Dosiahnuté výsledky môžu byť použité na podporu rozhodovania alebo vytváranie stratégie vo vnútri firmy, na podporu rozhodovania pri nákupe akcií alebo podielov vo firme, pri výbere partnerov a dodávateľov, atď.**

***Kľúčové slová* — Objavovanie znalostí, finančné dáta, predikcia bankrotu**

I. ÚVOD

Výkonnosť podniku a jeho finančná situácia závisí do veľkej miery od dosiahnutých výsledkov, preto je potrebné venovať analýze finančnej situácie značnú pozornosť. Túto analýzu je možné použiť nielen na hodnotenie súčasného stavu, ale aj na predpovedanie budúceho vývoja podniku. Dôležitým vstupným faktorom sú finančné dáta v potrebnej kvalite. V súčasnosti existuje mnoho databáz obsahujúcich obrovské množstvá finančných dát, ktoré sú však pre samotného človeka ťažko spracovateľné a pochopiteľne.

Jednou z možností ako z týchto dát získať potenciálne užitočné znalosti sú metódy objavovania znalostí. Proces objavovania znalostí je interaktívny a pozostáva z viacerých krokov. Pre dosiahnutie optimálnych výsledkov je pri tomto procese nevyhnutná asistencia človeka.

Tento článok je rozdelený do štyroch hlavných kapitol, kde prvá predstavuje krátky úvod do riešenej problematiky, druhá sa venuje možným spôsobom ako analyzovať finančnú situáciu firiem, tretia popisuje realizovaný proces objavovania na základe CRISP-DM a posledná zosumarizuje použitý postup a dosiahnuté výsledky.

II. ANALÝZA SÚČASNÉHO STAVU

A. Analýza finančnej situácie podniku

Finančná analýza predstavuje významnú súčasť podnikového riadenia. Patrí jej rozhodujúce postavenie pri hodnotení súčasného stavu a predpovedaní budúceho vývoja podniku. Finančná analýza umožňuje finančným analytikom ľahšie a presnejšie odhadnúť silné a slabé stránky podniku a problémy, ktoré vznikli, alebo môžu vzniknúť. [1]

Analýzu finančnej situácie podniku môžeme rozdeliť na [2]:

- Ex-post analýza – usiluje sa vysvetliť súčasnú finančno-ekonomickú situáciu podniku pohľadom do minulosti. Na dosiahnutých výsledkoch sa však už nemôže nič meniť.
- Ex-ante analýza – jej úlohou je predpovedať vývoj podnikových financií, prognózovať budúcu finančnú spôsobilosť podniku s presnosťou líšiacou sa podľa vzdialenosti cieľovej budúcnosti od súčasnosti.

V článku sa budeme venovať predovšetkým prognóze budúcej finančnej situácie podniku. V podnikovej praxi sa pri určení rizika odvetvia a podniku používajú rôzne typy modelov

predikčnej analýzy. Keďže poskytujú informácie o možnom nepriaznivom vývoji, často sú označované ako systémy včasného varovania.

B. Objavovanie znalostí

Objavovanie znalostí (OZ) predstavuje proces semi-automatickej extrakcie znalostí z databáz. Tieto znalosti musia byť v štatistickom zmysle platné, doposiaľ neznáme a potenciálne užitočné [3]. Proces objavovania znalostí je v rámci svojich jednotlivých krokov nedeterministický, s množstvom alternatívnych rozhodnutí. Preto výber akejkoľvek inej operácie, algoritmu, alebo zmena jeho parametrov vedie k iným výstupom, ktoré samozrejme ovplyvňujú ďalšie kroky. Človek rozhoduje o výbere vhodných operácií, algoritmov a ich parametrov za účelom objaviť skutočne cenné znalosti, ktoré môžu priniesť reálny prínos. Navyše len človek dokáže vyhodnotiť, ktorú z objavených znalostí je možné s úspechom aplikovať. [3]

C. CRISP-DM

Metodológia CRISP-DM (Cross Industry Standard Process for Data Mining) predstavuje referenčný model pre proces objavovanie znalostí a poskytuje prehľad o životnom cykle tohto procesu. Táto metodológia umožňuje proces objavovania znalostí urýchliť a zefektívniť, čo spôsobí zníženie nákladov. CRISP-DM pozostáva zo šiestich fáz, ktoré majú medzi sebou úzke vzťahy a obojstranné prechody: Pochopenie cieľa, Pochopenie dát, Príprava dát, Modelovanie, Vyhodnotenie a Nasadenie.

III. PREDIKCIA BANKROTU FIRIEM

Táto časť popisuje jednotlivé praktické kroky, operácie a experimenty s cieľom predikovať bankrot vybraných firiem na základe ich finančných dát pomocou podporného nástroja RapidMiner[1]. Všetky potrebné vstupné dáta boli získané z databázy európskych firiem Amadeus[2].

A. Pochopenie cieľa

Cieľom je predikovať finančnú situáciu firiem za určité časové obdobie, čiže či potenciálne zbankrotujú alebo nie. Z pohľadu objavovania znalostí ide o binárnu klasifikáciu do dvoch cieľových tried, kde cieľom je maximalizovať percento správne klasifikovaných prípadov.

Pre dosiahnutie tohto cieľa boli použité finančné dáta za roky 2003 až 2008 pôsobenia firiem. Je potrebné však zdôrazniť, že na ekonomickú situáciu firmy vplýva taktiež množstvo interných a externých faktorov, ktoré nie je možné pomocou týchto údajov zachytiť. Príkladom môže byť aj globálna ekonomická kríza, ktorá sa prejavila počas roka 2008. Tieto faktory je možné čiastočne eliminovať pomocou expertov na danú problematiku.

B. Pochopenie cieľa

Vstupné dáta boli získané z databázy európskych firiem Amadeus vo forme základných údajov o firmách a taktiež 31 ukazovateľov finančnej situácie. Vzhľadom na fakt, že veľkosť podniku v určitej miere pôsobí na výkonnosť firmy, a zároveň aj na jej finančnú a ekonomickú situáciu, sme sa rozhodli segmentovať výber dát z databázy na základe veľkosti firmy. Využili sme možnosť segmentácie ponúkanú databázou Amadeus, ktorá podniky rozdeľuje do štyroch kategórií:

- veľmi veľké spoločnosti (výnosy >= 100 mil. EUR, aktíva >= 200 mil. EUR, počet zamestnancov > 1000),
- veľké spoločnosti (výnosy >= 10 mil. EUR, aktíva >= 20 mil. EUR, počet zamestnancov > 150),
- stredné spoločnosti (výnosy >= 1 mil. EUR, aktíva >= 2 mil. EUR, počet zamestnancov > 15),
- malé spoločnosti (výnosy < 1 mil. EUR, aktíva < 2 mil. EUR, počet zamestnancov < 15).

Pre každú z týchto kategórií bol vykonaný osobitný export dát. Pre potrebu trénovacej množiny boli exportované podniky, ktoré v roku 2007 zaznamenali bankrot alebo aktívny stav. Spolu s týmito podnikmi boli exportované aj ich finančné údaje za obdobia rokov 2003 až 2007, ktoré predchádzali vzniknutej situácii. Konkrétne boli pre natrénovanie modelov použité ukazovatele za rok 2007, keďže práve v tomto roku boli vzhľadom na bankrot podnikov najväčšie rozdiely v hodnotách ukazovateľov aktívnych a bankrotujúcich podnikov. Následne

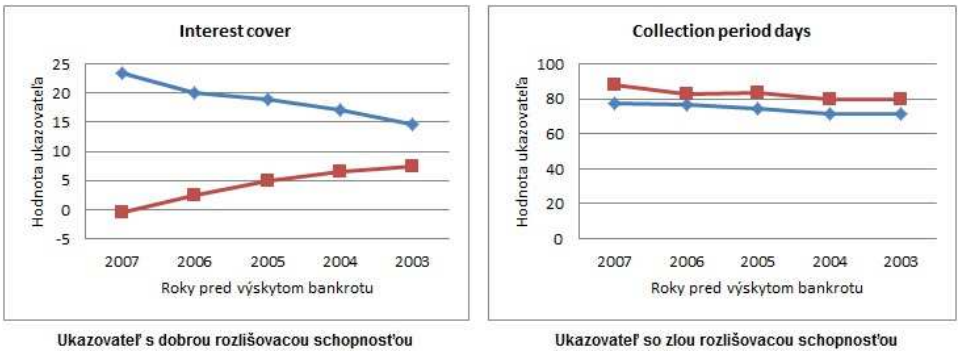
[1] http://rapid-i.com/content/view/181/190
[2] https://amadeus.bvdinfo.com/

sme si pre otestovanie modelov zvolili podniky, ktoré zaznamenali bankrot alebo aktívny stav v roku 2008. K týmto podnikom boli exportované ich finančné údaje za roky 2005 až 2008, kde sme vytvorený model otestovali pre každý rok osobitne. Cieľovým atribútom bol v prípade trénovacej množiny *Stav podniku* v roku 2007 a v testovacej množine *Stav podniku* v roku 2008.

C. Príprava dát

Výrazným problémom po načítaní dát bolo práve množstvo chýbajúcich záznamov. Tento problém sme sa snažili čiastočne eliminovať už pri exporte dát z databázy, keď sme eliminovali atribúty s príliš vysokým počtom chýbajúcich hodnôt. Taktiež sme nebrali do úvahy ani všetky základné údaje o podniku, keďže jeho finančnú situáciu chceme zistiť pomocou finančných ukazovateľov.

Po týchto úpravách sme získali dátovú množinu, ktorá obsahovala 20 ukazovateľov a jeden cieľový atribút *Stav podniku*. Tieto dáta boli ďalej použité napr. na výber ukazovateľov s dobrou rozlišovacou schopnosťou. Tento výber bol vykonaný na základe grafického porovnania priemerných hodnôt ukazovateľov v časovom rade. Po tomto grafickom porovnaní sme ignorovali 8 ukazovateľov vzhľadom na ich zlú rozlišovaciu schopnosť. Príklad ukazovateľa s dobrou a ukazovateľa so zlou rozlišovacou schopnosťou je znázornený na obrázku, viď. Obr.1.



Obr. 1 Rozlíšenie ukazovateľov pomocou grafického porovnania priemerných hodnôt

Ako druhý krok sme pre zvyšných 12 ukazovateľov vykonali test nezávislosti pomocou korelačnej tabuľky. Na základe vytvorenej korelačnej tabuľky bolo zvolených 5 ukazovateľov s najvyššou mierou nezávislosti. Týchto 5 ukazovateľov tvorilo základ pre tvorbu modelov. Ide konkrétne o tieto ukazovatele: *Úrokové krytie*, *Bežná likvidita*, *Obrat aktív*, *Rentabilita vlastného kapitálu* a *Stupeň samofinancovania*. Tieto ukazovatele tvorili základ vytvorených klasifikačných modelov.

Poslednou operáciou predspracovania bola integrácia oboch dátových množín a to v prvom prípade záznamy popisujúce bankrotujúce podniky a na druhej strane aktívne pokračujúce. Keďže záznamy s bankrotujúcimi podnikmi boli výrazne menšie ako záznamy s aktívnymi, bolo potrebné tento nepomer vybalansovať. Tento problém sme odstránili zúžením výberového súboru, viď. Tab.1.

Tabuľka 1
Počet záznamov po zúžení výberového súboru

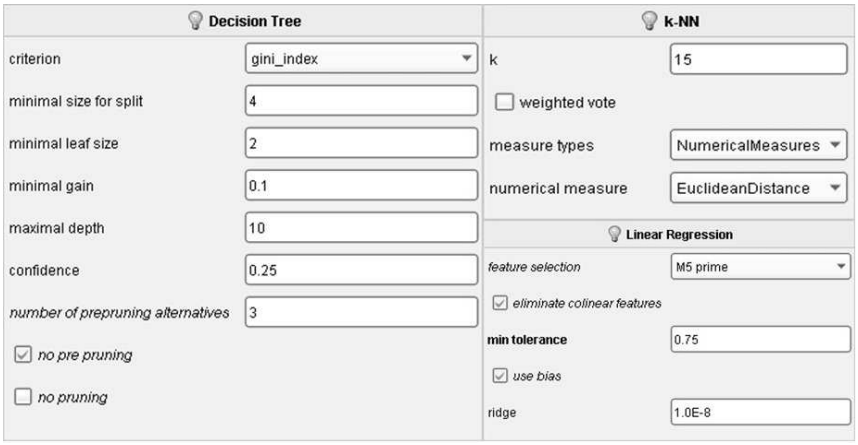
Kategória podnikov	Trénovacia množina		
	Aktívne	Bankrotujúce	Spolu
Malé	540	540	1080
Stredné	1000	1000	2000
Veľké	250	250	500

D. Modelovanie

Vo fáze modelovanie sme sa venovali výberu vhodných algoritmov a ich vstupných parametrov. Na vytvorenie klasifikačných modelov boli použité tri algoritmy a to rozhodovacie stromy, lineárna regresia a k-najbližších susedov. Tieto modely boli aplikované v rámci

všetkých kategórií podnikov.

Trénovacia množina obsahovala údaje o firmách, ktoré zaznamenali bankrot alebo aktívny stav v roku 2007 a testovacia množina obsahuje údaje o firmách, ktoré tento stav zaznamenali v roku 2008. Keďže naším cieľom bolo predpovedať stav podniku 2 roky, 1 rok pred a v tom istom roku ako nastal, tak pre testovanie boli použité tie isté ukazovatele avšak postupne za roky 2006, 2007 a 2008. Tvorba modelov mala iteratívny charakter, kde sme pri trénovaní modelov menili nastavenia príslušných parametrov pre jednotlivé algoritmy vzhľadom na ich úspešnosť klasifikácie.



Obr. 2 Nastavenie parametrov pre jednotlivé algoritmy v prostredí RapidMiner

E. Vyhodnotenie

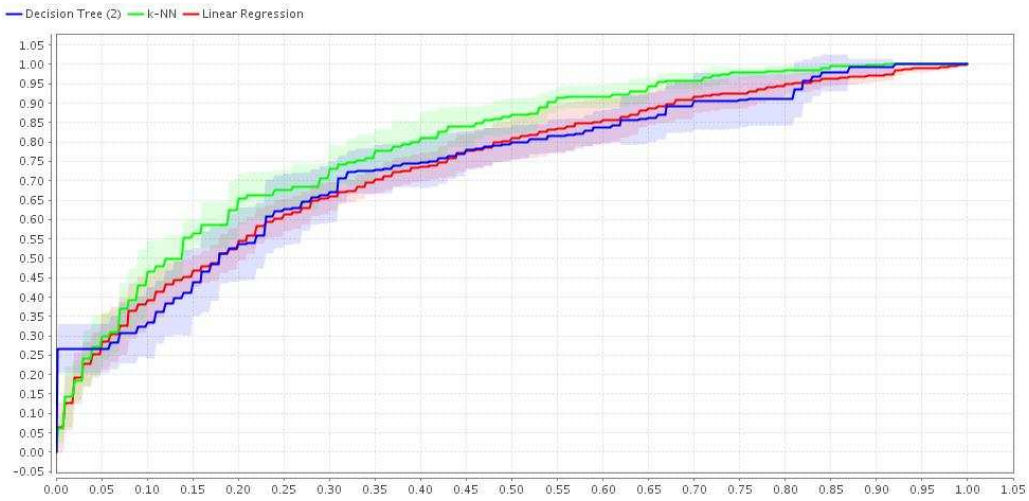
Na vyhodnotenie kvality vytvorených modelov bola použitá testovacia množina. Percentuálne najvyššia úspešnosť klasifikácie bola zaznamenaná pri použití modelu k-NN a to pre všetky kategórie podnikov. Svedčí o tom aj fakt, že v čase bankrotu firiem sa vo všetkých kategóriách úspešnosť pohybovala okolo hranice 70%. Najvyššiu úspešnosť v tomto prípade dosiahol model pri stredných podnikoch a to 71,78% pre rok, v ktorom bol zaznamenaný bankrot, viď. Tab.2.

Tabuľka 2
Úspešnosť modelu k-NN pre jednotlivé kategórie podnikov

Kategória podniku	Časové obdobie		
	rok bankrotu	1 rok pred bankrotom	2 roky pred bankrotom
Malé podniky	68,88%	66,21%	63,61%
Stredné podniky	71,78%	63,46%	61,50%
Veľké podniky	68,00%	71,40%	66,96%

Zaujímavým zistením je, že percentuálna úspešnosť klasifikácie v kategórii malých a stredných podnikov s blížiacim sa bankrotom rástla, avšak v prípade veľkých podnikov bola úspešnosť 1 rok pred bankrotom väčšia ako v roku zaznamenania bankrotu. Presnosť modelov je možné porovnať aj graficky pomocou ROC kriviek, viď. Obr.3.

Pri návrhu experimentoch sme zvolili segmentáciu podľa veľkosti podnikov. Ďalšou možnosťou je segmentácia podnikov na základe odvetvia, v ktorom pôsobia. Táto segmentácia sa javí ako rozumnejšie riešenie, keďže jednotlivé odvetvia sa líšia svojim pôsobením, čo sa prejaví aj na dosiahnutých hodnotách ukazovateľov. V našich experimentoch sme uvažovali aj nad segmentáciou podľa odvetvia, avšak vzhľadom na pomerne nízky počet záznamov sme od tohto úmyslu ustúpili.



Obr. 3 ROC krivky výsledných modelov pre kategóriu stredných podnikov

IV. ZÁVER

Tento článok popisuje použitie troch algoritmov dolovania dát pre predikciu finančnej situácie firiem (bankrotujúca alebo aktívna) na základe dát z finančnej databázy Amadeus a využitím voľne dostupného prostredia RapidMiner. Výsledné klasifikačné modely predstavujú zaujímavú alternatívu pre získanie dôležitej informácie pre podporu rozhodovania manažérov firiem, potenciálnych kupcov alebo partnerov.

V rámci realizovaného procesu objavovania znalostí sme zaznamenali niekoľko problémov, ktoré sme sa čiastočne snažili odstrániť. Výrazným problémom bola nekompletnosť vstupných dát a množstvo chýbajúcich údajov pri jednotlivých ukazovateľoch, čo považujeme ako zásadnú nevýhodu použitej databázy Amadeus. Po ich eliminácii z ďalšieho postupu sa nám výrazne znížil počet záznamov dátovej množiny, čo sa prejavilo aj na presnosti vytvorených modelov.

Celková úspešnosť klasifikácie takisto silno závisí aj od spôsobu výberu ukazovateľov pre tvorbu modelov. V tejto oblasti existujú aj iné prístupy na výber vhodnej množiny vstupných ukazovateľov pre hodnotenie výkonnosti podniku, ktoré môžu byť doplnené aj o znalosti expertov. Takéto znalosti môžu prispieť k zvýšeniu presnosti klasifikačných modelov.

POĎAKOVANIE

Tento príspevok vznikol za podpory KEGA projektu č. 065TUKE-4/2011 (50%) a VEGA projektu č. 1/1147/12 (50%).

LITERATÚRA

[1] R. Kotulič, P. Király, M. Rajčániová, *Finančná analýza podniku*, druhé vydanie. Bratislava: Iura Edition, spol. s. r. o., 2010.
[2] K. Zalai a kol., *Finančno-ekonomická analýza podniku*, piate vydanie. Bratislava: Sprint, 2007.
[3] J. Paralič, *Objavovanie znalostí v databázach*. Košice: Elfa, 2003.
[4] RAPID-I Report the future: RapidMiner. Dostupné na internete: <http://rapid-i.com/content/view/181/190/lang,en/>

Využitie BI na analýzu finančnej situácie firiem

Cecília HAVRILOVÁ, František BABIČ

Katedra kybernetiky a umelej inteligencie, Fakulta elektrotechniky a informatiky,
Technická univerzita v Košiciach, Slovenská republika

cecilia.havrilova@student.tuke.sk, frantisek.babic@tuke.sk

***Abstrakt* — Business Intelligence (BI) a jej nástroje poskytujú zaujímavé možnosti pre analýzu finančnej situácie firiem. Prostredníctvom vytvoreného dátového skladu a analytických funkcií je možné získať rôzne pohľady na zhromaždené dáta, ktoré môžu byť ďalej použité na podporu rozhodovania, definovanie stratégie, riešenie krízových situácií, atď. Tento príspevok je venovaný analýze firiem na základe dát z dvoch databáz Albertina a Creditinfo, poskytujúcich nielen všeobecné informácie o firmách, ale aj príslušné finančné informácie získané prostredníctvom účtovných uzávierok. Praktické experimenty boli realizované v troch vybraných nástrojoch BI: VanillaBI, SpagoBI a Pentaho, z ktorých práve posledné riešenie bolo vyhodnotené ako najvhodnejšie z používateľského hľadiska.**

***Kľúčové slová* — Business Intelligence, Dátový sklad, OLAP**

I. Úvod

Oblasť informačných systémov je v súčasnosti predmetom veľkého záujmu akademickej ako aj komerčnej sféry. Podniky a organizácie si čoraz viac uvedomujú strategickú hodnotu informácií uložených vo svojich prevádzkových informačných systémoch. V dôsledku toho investujú množstvo prostriedkov do budovania jednotného úložiska dát, nad ktorým by bolo možné - za pomoci súčasných BI riešení, tieto dáta analyzovať, získavať z nich dôležité informácie a tým dosiahnuť významnú konkurenčnú výhodu.

Tento príspevok je rozdelený do piatich hlavných častí, kde prvá predstaví v krátkosti riešenú problematiku a základné teoretické východiská, druhá a tretia kapitola sú venované konkrétnym krokom implementácie, vo štvrtej kapitole sú načrtnuté ukážky použitých analýz a posledná kapitola sumarizuje článok a jeho výsledky.

A. Business Intelligence

BI riešenia predstavujú rozsiahlu sadu konceptov a metód určených pre skvalitnenie rozhodovacích procesov v organizáciách prostredníctvom analýzy dát, reportovania a dopytovacích nástrojov [1]. Riešenia BI pozostávajú z niekoľkých vrstiev komponentov, napr. Vrstva pre extrakciu, transformáciu, čistenie a nahrávanie dát; Vrstva zabezpečujúca správu dát a ich ukladanie do dátových skladov a úložísk; Vrstva pre analýzu dát prostredníctvom dolovania, resp. OLAP operácií; Prezentačná vrstva určená pre koncových používateľov a pod.

B. Dátový sklad

Dátový sklad je základným stavebným prvkom BI riešení. Je to miesto, v ktorom sa uchovávajú dáta získané z interných ako aj externých zdrojov. Tie sú transformované do vhodnej podoby tak, aby ich bolo možné analyzovať a zobrazovať koncovým používateľom v rôznych výstupných formátoch. Vo všeobecnosti môžeme povedať, že ide o databázu reprezentujúcu dáta, modelovanú vo forme multidimenzionálnej dátovej kocky, ktorá sa skladá z *dimenzií* a *faktov*.

Dimenzie predstavujú kolekcie členov rovnakého typu pohľadu. Každá dimenzia sa skladá z jedného alebo viacerých atribútov, pričom môže vytvárať logicky alebo organizačne usporiadanú hierarchiu.

Fakty sú konkrétne numerické hodnoty alebo javy pričom každá hodnota sa nachádza v priesečníku najnižších úrovní vybraných dimenzií - bunke. [2]

II. NÁVRH DÁTOVÉHO SKLADU

A. Popis zdrojových dát

Pre tvorbu dátového skladu boli použité zdrojové dáta extrahované z dvoch komerčných databáz slúžiacich k získavaniu informácií o registrovaných firmách a neziskových organizáciách v Slovenskej a Českej republike. Ide o databázy:

- *ALBERTINA* – súbor databáz a služieb všetkých registrovaných firiem a organizácií na území SR a ČR, ktorým bolo pridelené IČO.
- *CREDITINFO* – databáza zameraná na poskytovanie finančných informácií o firmách, získaných prostredníctvom účtovných závierok.

Obe databázy spolu obsahovali údaje o viac ako 600 000 firmách a organizáciách na území SR, pričom len pri 10 313 subjektoch boli k dispozícii aj účtovné závierky. V rámci týchto firiem však boli zaznamenané aj viaceré chýbajúce údaje potrebné pre analýzy. V dôsledku toho bola pre potreby tvorby dátového skladu vybraná vzorka firiem s kompletnými údajmi o veľkosti 3 337 subjektov a k nim prislúchajúcich 12 613 účtovných závierok. Celkovo obsahovala zdrojová databáza 7 tabuliek vyexportovaných vo formáte MDB, vzájomne prepojených prostredníctvom IČO. (Tab. 1)

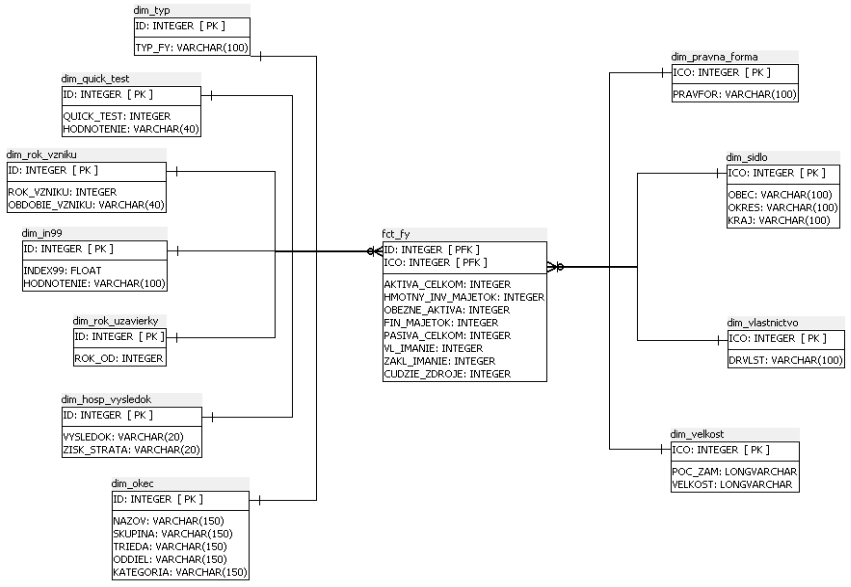
Tabuľka 1
Zdrojové dátové tabuľky

Tabuľka	Počet atribútov	Počet záznamov	Obsah tabuľky - údaje
FIRMY	79	3 337	Identifikačné, Kontaktné, Personálne, Registračné, Obchodné
FIRDCER	8	396	Adresa dcérskej spoločnosti, % podiel mat. spoločnosti
FIRPOB	10	6 225	Kontaktné údaje o pobočkách firiem
FIRVLS	11	4 091	Informácie o vlastníkoch firiem, ich vklad a podiel vo firme
OSOBA	13	15 895	Mená kontaktných osôb, ich funkcia, informácia o rezorte
UCTYDBF	231	12 613	Finančné údaje o firmách prostredníctvom účtovných závierok
UCTYPF	190	12 613	Vybrané finančné údaje z *UCTYDBF* doplnené o ekonomické indexy

B. Návrh štruktúry dátového skladu

Pre návrh dátového skladu bola použitá hviezdicová štruktúra tvorená jednou tabuľkou faktov a 11 tabuľkami dimenzií. Tabuľku faktov tvoria atribúty *IČO, celkové aktíva, celkové pasíva, hmotný investičný majetok, obežné aktíva, finančný majetok, vlastné imanie, základné imanie a cudzie zdroje.*

Tabuľky dimenzií popisujú firmy z hľadiska: *právnej formy, sídla, vlastníctva, veľkosti, typu firmy, odvetvovej klasifikácie ekonomickej činnosti, podľa obdobia vzniku a roku zverejnených účtovných závierok a podľa ekonomických ukazovateľov – hospodársky výsledok, quick test a index* IN_{99}. (Obr. 1)



Obr. 1 Návrh štruktúry dátového skladu.

III. IMPLEMENTÁCIA

Tieto dátové tabuľky boli spracované pomocou MySQL databázového servera, ktorý slúžil aj ako centrálne úložisko pre tvorbu tabuliek dimenzií a tabuľky faktov. Analýza dát nad vytvoreným dátovým skladom prebiehala prostredníctvom troch zvolených BI riešení: VanillaBI, SpagoBI a PentahoBI.

A. VanillaBI

Balík riešenia pozostáva z VanillaBI Servera dostupného vo verzii nakonfigurovanej pre MySQL. Riešenie ponúka aj niekoľko samostatných modulov určených pre správu používateľov a správu dát. Vo fáze ETL bol použitý modul Vanilla BIGateWay umožňujúci spojenie s MySQL zdrojovými tabuľkami a následnú transformáciu dát do tabuliek dimenzií a tabuľky faktov. Pre overenie potenciálu daného BI riešenia boli vykonané OLAP analýzy nad vytvoreným dátovým skladom v prostredí modulu Free Analysis Schema Designer. Samotným analýzam predchádzala tvorba dátovej kocky, ktorá vznikla vzájomným prepojením tabuľky dimenzií a tabuliek faktov prostredníctvom ich kľúčov. Pre tvorbu reportov bol použitý modul Vanilla Web Report, doplnený o komponent BIRT.

B. SpagoBI

Riešenie SpagoBI pozostáva z SpagoBI Server a modulov pre správu dát. Na integráciu a správu dát vo fáze ETL bol použitý nástroj Talend Open Studio. V zaujímavom a používateľsky atraktívnom grafickom prostredí boli vykonané operácie transformácie dát do požadovaných tabuliek dimenzií a tabuľky faktov. Na tvorbu biznis modelu dátovej kocky bol použitý nástroj SpagoBI Studio umožňujúci získavanie výsledkov vo veľmi krátkom čase pre konkrétne dopyty. Pre rozsiahle OLAP operácie bola vytvorená XML schéma v prostredí nástroja Mondrian Workbench Schema. Tvorba reportov prebiehala v systéme iReport.

C. PentahoBI

Balík riešenia PentahoBI pozostáva z niekoľkých aplikácií, pričom každá z nich rieši určitý proces v rámci BI – implementáciu dátového skladu, vykonávanie analýz a tvorbu reportov. Transformácia dát prebiehala v samostatnom nástroji Pentaho Data Integration s výstupom vo forme MySQL tabuliek dimenzií a tabuľky faktov. Tieto boli načítané do prostredia Pentaho BI Server, v ktorom boli nad vytvoreným dátovým skladom vykonané rôzne OLAP analýzy. Ako reportovací nástroj bol zvolený integrovaný modul v rámci Pentaho BI Server umožňujúci tvorbu ad-hoc reportov.

D. Porovnanie BI riešení

Predstavené voľne dostupné BI riešenia sú v súčasnosti silnými konkurentmi. Každé z týchto riešení poskytuje výkonné moduly pre správu dát, tvorbu dátového skladu a realizovanie analýz a reportov.

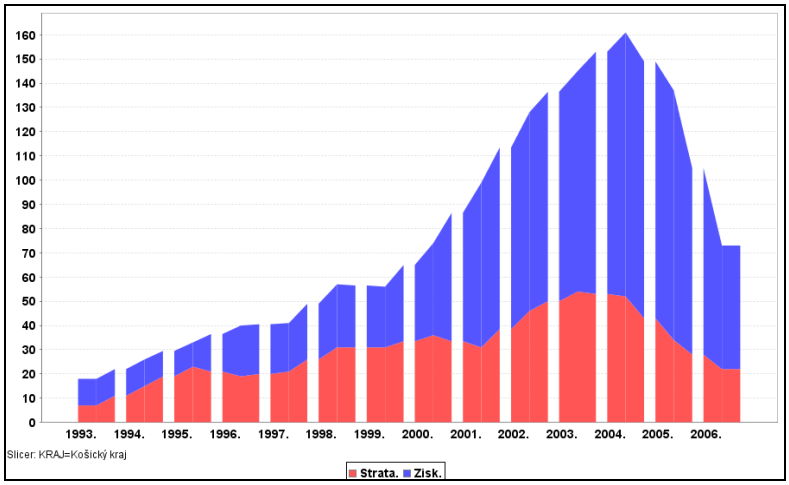
Na základe vykonaných experimentov a v rámci porovnania funkcionalít vybraných riešení, bol pre tvorbu rozsiahlejších analýz zvolený softvér PentahoBI. Dôvodom je najmä skutočnosť, že tento softvér je komplexným riešením s integrovanými nástrojmi pre vytváranie dátového skladu, analýz a reportov priamo na jednom mieste.

IV. ANALÝZA FINANČNEJ SITUÁCIE FIRIEM

Nad dátovým skladom je možné pomocou analytických funkcií vykonávať rôzne typy analýz, v tomto prípade boli použité tri významné ekonomické ukazovatele. Prvé dva ukazovatele - *index* IN_{99} a *Quick test*, hodnotia finančnú kondíciu a finančné zdravie podniku. Posledný ukazovateľ - *Hospodársky výsledok účtovného obdobia*, meria výsledok hospodárenia podniku.

Analýza uvedená v tejto časti článku je zameraná na popis firiem pôsobiacich v Košickom kraji na základe ekonomického ukazovateľa *Hospodársky výsledok účtovného obdobia*.

Z dostupných účtovných závierok pre Košický kraj, v počte 1 089 za roky 1993-2006 vyplýva, že až 61% firiem vykazovalo v danom období zisk. Zvyšných 39% firiem uvádza v danom období stratu. V rokoch 1994 - 1999 prevláda v dostupných účtovných závierkach počet firiem vykazujúcich stratu nad počtom firiem vykazujúcich zisk. V roku 2000 sa ale tento trend mení – v účtovných závierkach prevládajú hospodárske výsledky vo forme zisku nad stratovými výsledkami. (viď. Obr. 1)



Obr. 2 Hospodárske výsledky 1993-2006, Košický kraj.

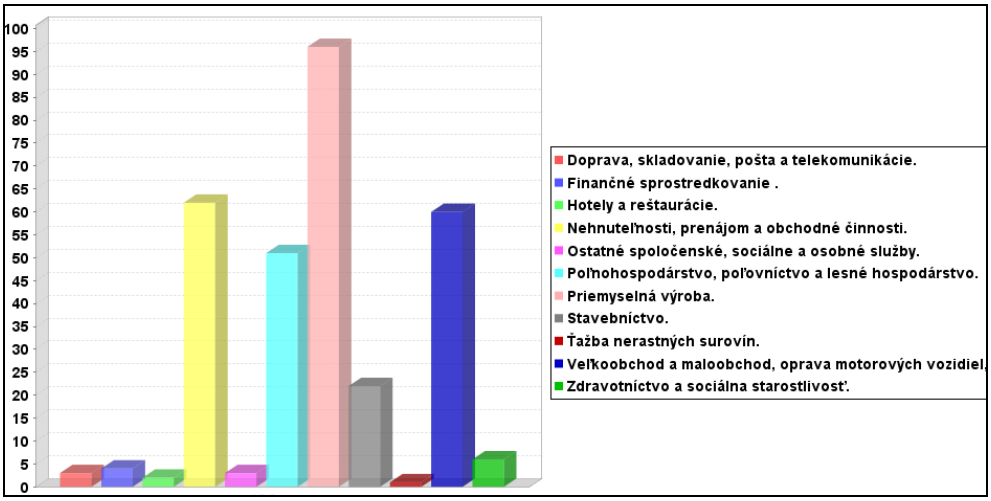
Použitím funkcie filtrovania boli ďalšie analýzy aplikované na dáta, v ktorých *Hospodársky výsledok* bol vo forme *Straty*. Prostredníctvom *drill down operácie* bol získaný aj detailnejší pohľad na dáta. Tento pohľad poukazuje na účtovné závierky rozdelené do troch kategórií:

- *Strata do 100 tis.* – 11,14 % účtovných závierok
- *Strata do 500 tis.*- 15,40 % účtovných závierok
- *Strata nad 500 tis.* – 73,46 % účtovných závierok

Keďže najväčší podiel na závierkach s *Hospodárskym výsledkom vo forme straty* tvoria závierky so Stratou nad 500 tis., práve tie sa stali objektom ďalšieho skúmania.

Pridaním dimenzie *dim_okec* bol získaný pohľad na odvetvovú klasifikáciu ekonomickej činnosti uvedenú v účtovných závierkach. (viď. Obr. 2) Najviac zastúpené odvetvia sú:

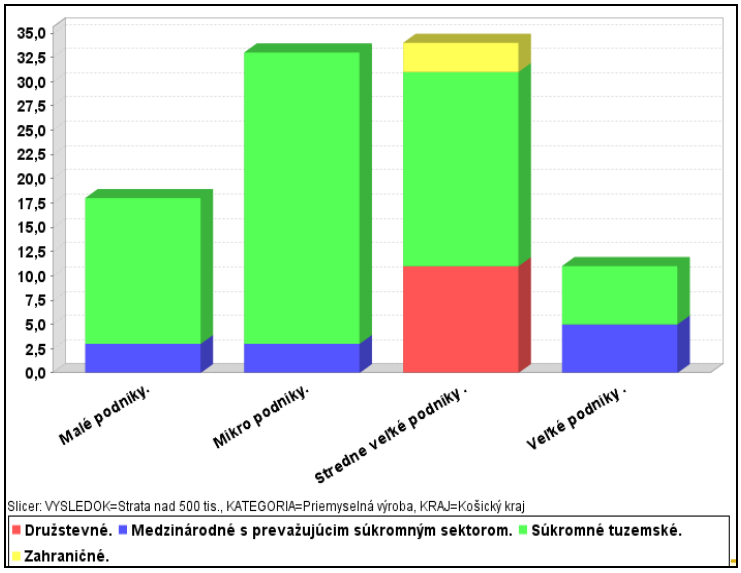
- *Priemyselná výroba* – 30,96% účtovných závierok
- *Nehnuteľnosti, prenájom a obchodné činnosti* – 20% účtovných závierok
- *Veľkoobchod a maloobchod, oprava motorových vozidiel, motocyklov a spotrebného tovaru* – 19,35% účtovných závierok
- *Poľnohospodárstvo, poľovníctvo a lesné hospodárstvo* – 16,45% závierok



Obr. 3 Klasifikácia ekonomickej činnosti podľa odvetví, Košický kraj.

Nastavením filtra pre *dim_okec* bolo pre ďalšie potreby skúmania zvolené odvetvie *Priemyselná výroba.* V tomto odvetví a za daných podmienok patrí takmer 74% závierok *Súkromným tuzemským firmám.* Viac ako 11% závierok patrí *Družstevným firmám* a *Medzinárodným firmám s prevažujúcim súkromným sektorom.* Najmenší počet závierok patrí *Zahraničným firmám.*

Poslednou operáciou - *drill down*, bola pridaná dimenzia *dim_velkost.* Z tohto pohľadu patrí najväčšia časť závierok *Mikro podnikom* (34,75%) a *Stredne veľkým podnikom* (35,41%), pričom u *Stredne veľkých podnikov* možno badať najväčšiu rôznorodosť vlastníckych vzťahov.



Obr. 4 Rozdelenie podnikov podľa veľkosti, Košický kraj.

Prostredníctvom OLAP operácií je možné vykonávať ďalšie analýzy firiem prostredníctvom iných ekonomických ukazovateľov, prípadne analyzovať finančnú situáciu konkrétnych firiem pomocou ich účtovných závierok.

V prípade analýzy konkrétnych firiem sú k dispozícii aj ďalšie atribúty tabuľky faktov - *celkové aktíva, celkové pasíva, hmotný investičný majetok, obežné aktíva, finančný majetok, vlastné imanie, základné imanie a cudzie zdroje.* Vývoj týchto atribútov je možné sledovať v časovom rozmedzí zverejnených účtovných závierok.

V. Záver

Schopnosť premieňať dáta, získané prostredníctvom každodenných transakcií na hodnotné informácie, vedie k významným zlepšeniam v oblasti riadenia a rozhodovania. Pre správne taktické ale aj strategické rozhodnutia nie je nutné mať k dispozícii veľké množstvo dát, dôležité je ale vedieť analyzovať dostupné dáta a odvodiť z nich nové a potrebné znalosti.

Tento príspevok popisuje návrh a implementáciu budovania dátového skladu v prostredí troch vybraných BI riešení. Pre demonštráciu potenciálu zvoleného nástroja PentahoBI bola vykonaná všeobecná analýza firiem sídliacich v Košickom kraji. Táto analýza vychádza zo zverejnených účtovných závierok firiem, ktorých hospodársky výsledok bol v sledovanom období vo forme straty. Analýzou bolo preukázané, že najviac takýchto firiem sa nachádza v sektore *Priemyselnej výroby*. Ide pritom o *Mikro podniky* a *Stredne veľké podniky*, ktoré sú v *Súkromnom vlastníctve*.

Implementované BI riešenie dáva používateľom možnosť vykonávať rôzne OLAP operácie v reálnom čase prostredníctvom webového rozhrania. V budúcnosti je možné rozšíriť súčasný dátový sklad o aktualizované dáta, prípadne implementovať ďalšie dátové sklady, nad ktorými by bolo možné vykonávať nové analýzy.

Poďakovanie

Tento príspevok vznikol za podpory KEGA projektu č. 065TUKE-4/2011 (50%) a VEGA projektu č. 1/1147/12 (50%).

Literatúra

[1] O. Novotný, J. Pour, D. Slánský, *Business Intelligence: Jak využít bohatství ve vašich datech*. Grada Publishing a.s., 2005.
[2] V. Rainardi, *Building a data warehouse with examples in SQL Server*. Apress, 2008.

Využitie neurónovej siete FALCON pre riadenie pohybu robota Nao pri kope

[1]*Dušan HODOŠKO*, [2]*Ján VAŠČÁK*

[1]Centrum inteligentných technológií
[2]Katedra kybernetiky a umelej inteligencie, Fakulta elektrotechniky a informatiky,
Technická univerzita v Košiciach, Slovenská republika

[1]dusan.hodosko@gmail.com, [2]jan.vascak@tuke.sk

***Abstrakt* — Cieľom tejto práce je vytvorenie riadiaceho systému vyhľadávania a orientácie robota Nao pre potreby kopu v robofutbale a využitie neurónovej siete FALCON v tomto systéme. Riadiaci systém robota má byť schopný identifikovať loptu a bránu na ihrisku, sledovať loptu chôdzou a natočiť robota do polohy tak, aby sa pred ním nachádzala lopta a brána. Neurónová sieť FALCON je použitá na natáčanie robota okolo lopty. Riadiaci systém je implementovaný v aplikácii, ktorá beží v počítači a komunikuje s robotom cez Wi-Fi.**

***Kľúčové slová* — neurónová sieť, fuzzy regulátor, FALCON, robofutbal, Nao, robot**

I. ÚVOD

Každý rok sa konajú majstrovstvá sveta v robofutbale (RoboCup), ktoré sa hrajú s humanoidnými robotmi Nao. Základnou vlastnosťou RoboCup-u je, že každý robot je samostatný hráč, ktorý rozhoduje sám za seba.

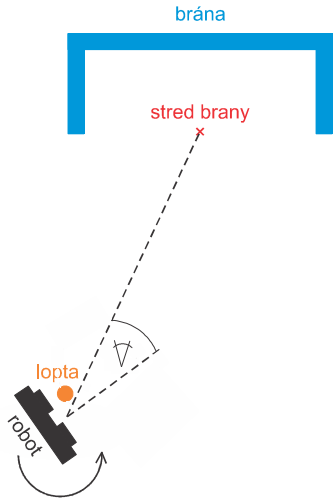
Našou ambíciou bolo vytvorenie robofutbalu, kde by roboti boli ovládaní jediným počítačom. Tento počítač sa dá chápať ako tréner (coach), ktorý bude ovládať pohyb a stratégiu robotického mužstva. Keďže sa jedná o veľmi rozsiahly problém a na jeho realizáciu by bol potrebný tím viacerých ľudí, zúžili sme problematiku na ovládanie jedného robota, ktorý dokáže chodiť za loptou a v prípade, že je pri lopte a vidí bránu, tak sa dokáže správne otáčať okolo lopty, až kým nie je na pozícii, z ktorej by mohol vystreliť na bránu. Neurónová sieť FALCON bola využitá pri natáčaní robota okolo lopty. Navrhnutý riadiaci systém je naprogramovaný v jazyku C#.

II. RIADIACI SYSTÉM ROBOTA NAO

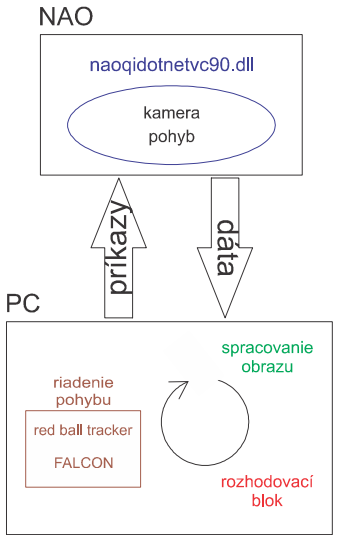
Riadiaci systém robota Nao je navrhnutý tak, aby spĺňal požiadavku, že robot má byť ovládaný cez Wi-Fi pomocou počítača. Hlavnou funkciou riadiaceho systému je sledovanie lopty chôdzou, správne natočenie robota na bránu a v konečnej fáze kopnutie lopty na bránu. Pri sledovaní lopty chôdzou môže lopta stáť na jednom mieste alebo byť v pohybe. V prípade, že sa robot priblíži k lopte na takú vzdialenosť, že je schopný kopnúť do lopty, začne hľadať bránu. Po nájdení brány robot určí približný uhol natočenia brány vzhľadom na jeho pozíciu a bude sa snažiť otáčať okolo lopty až do chvíle, kým lopta a brána nebudú priamo pred ním respektíve, kým uhol natočenia brány nebude približne nulový (Obr. 1). Po tejto fáze nasleduje samotný kop.

Podľa uhla natočenia brány sa určí smer otáčania robota okolo lopty. V prípade, že je uhol natočenia brány záporný, robot sa začne otáčať do pravej strany (Obr. 1). V opačnom prípade sa bude otáčať do ľavej strany. Na otáčanie okolo lopty je použitá neurónová sieť FALCON s dvoma vstupmi a dvoma výstupmi. Vstupmi do neurónovej siete sú súradnice polohy lopty na obraze a výstupmi sú parametre pre chôdzu. Podrobnejší popis riadiaceho systému bude vysvetlený v nasledujúcich kapitolách.

Tento riadiaci systém sa skladá z viacerých blokov, pričom ho môžeme rozdeliť na dve základné časti a to na riadiacu časť (PC) a vykonávaciu časť (Nao) (Obr. 2). Z robota sa získava obraz a uhly natočenia kĺbov. Na základe týchto dát program v počítači rozhoduje čo má Nao spraviť. Dá sa povedať, že Nao je v podstate externé zariadenie na zber dát, ktoré dostáva príkazy z počítača a následne ich vykonáva.



Obr. 1 Určenie uhla natočenia brány



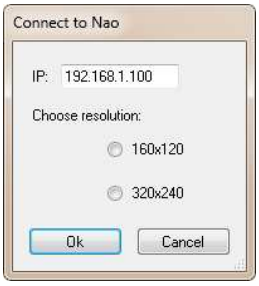
Obr. 2 Štruktúra riadiaceho systému

Spoločnosť Aldebaran, ktorá vyrába robota Nao, vytvorila knižnicu s názvom naoqidotnetvc90.dll, pomocou ktorej je možné programovať robota v jazyku C#. Tento jazyk som si vybral pretože je objektový a je vhodný pre väčšie projekty, pričom rýchlosť výpočtov je dostačujúca, keďže základné funkcie, ktoré sú uložené v robotovi sú naprogramované v programovacom jazyku C++. Pomocou knižnice naoqidotnetvc90.dll dokáže riadiaci systém robota, ktorý beží na desktopovom počítači a je naprogramovaný v jazyku C#, komunikovať s týmito funkciami. Tento riadiaci systém vyhľadávania a orientácie riadi robota a spracováva dáta v reálnom čase. V nasledujúcej časti budú popísané jednotlivé časti riadiaceho systému (Obr. 2).

III. SPRACOVANIE OBRAZU

Obraz z robota sa získava pomocou funkcie getImageRemote, ktorá vracia návratovú hodnotu objekt, v ktorom sa nachádza pole bajtov. Táto funkcia sa nachádza v spomínanej knižnici naoqidotnetvc90.dll. Získané pole bajtov obsahuje pre každý pixel obrazu stále trojicu hodnôt RGB v rozmedzí od 0 do 255. Najskôr sa pole bajtov transformuje na objekt triedy Bitmap, ktorá sa nachádza v štandardnom SDK jazyka C#. Ďalej sa už pracuje s týmto objektom triedy Bitmap. Tento proces sa opakuje niekoľko krát za sekundu (FPS), čím získame video v reálnom čase. Rýchlosť zobrazovania obrazu je nízka, pretože komunikácia s robotom je realizovaná cez Wi-Fi. Najnovšia verzia Naa dokáže s rozlíšením 160x120 pixlov posielať obraz 10 až 15 krát za sekundu a na rozlíšení 320x240 pixlov 2 až 5 krát za sekundu. Dá sa nastaviť aj rozlíšenie 640x480 pixlov, ale FPS je veľmi nízke, preto je toto rozlíšenie vhodné najmä ak chceme pracovať s obrázkami. Pre aplikácie bežiace v reálnom čase je lepšie používať menšie rozlíšenie 160x120 pixlov kvôli vyššiemu FPS. V mojom riadiacom systéme je možné pracovať s

rozlíšením 160x120 a 320x240 pixlov. Veľkosť rozlíšenia si užívateľ nastaví hneď pri pripájaní sa na robota, kde musí zadať IP robota a veľkosť rozlíšenia (Obr. 3).



Obr. 3 Výber rozlíšenia

A. Prahovanie

Prahovanie je podľa [2] najjednoduchší spôsob segmentácie obrazu a z pohľadu výpočtovej náročnosti je najrýchlejšou segmentačnou metódou. Vychádza z predpokladu, že objekty a im zodpovedajúce oblasti na obraze sú charakterizované približne rovnakou jasovou úrovňou. Ak obraz obsahuje dostatočne kontrastné objekty vzhľadom na pozadie, je možné použiť určitú jasovú úroveň, tzv. prah k ich vzájomnému oddeleniu. Prahovanie je podľa [2] definované nasledovne

$$f(i,j) = d,\ ak\ t_{d-1} \le f(i,j) < t_d;\ d = 1,2,...,D \tag{1}$$

kde t_d je d-tý prah a *f(i,j)* je segmentovaná hodnota v bode obrazu *(i,j)*.

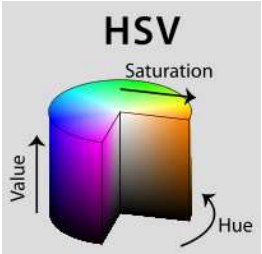
Pri spracovávaní obrazu sme použili prahovanie, keďže lopta aj brána má svoju špecifickú farbu, ktorá je dopredu známa a povrch ihriska je okrem bielych čiar celý zelený. Existuje viacero spôsobov ako vyberať prah. Pri hraní robofutbalu sa môžu často meniť svetelné podmienky a preto sme vybrali interaktívny spôsob výberu prahu. V programe jednoducho pozastavíme video z robota stlačením tlačidla pauza a označíme časť obrazu myšou, ktorý chceme prahovať (Obr. 4).



Obr. 4 Video z robota s rozlíšením 320x240 pixlov

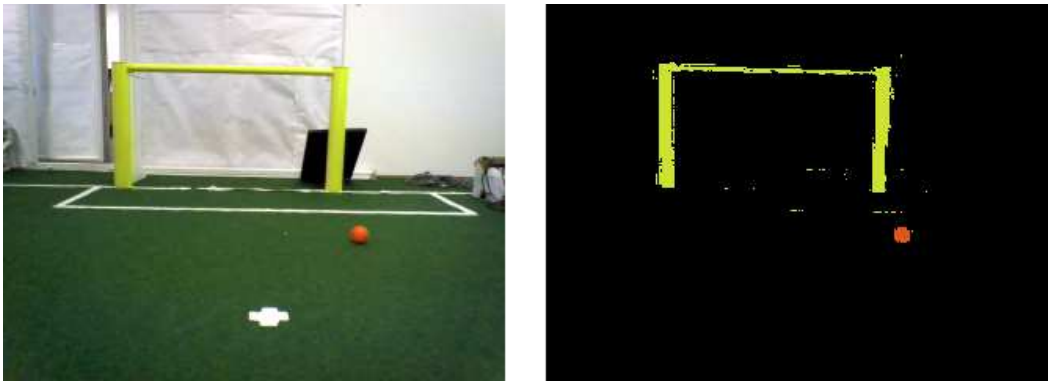
Z označenej časti obrazu sa zoberú všetky pixely a prah sa určí ako aritmetický priemer jasov týchto pixlov. Po stlačení tlačidla hraj je video z robota prahované na základe nastaveného prahu. Tento spôsob sa osvedčil, pretože je veľmi dynamický a umožňuje rýchle zmenenie prahu počas behu programu.

Pri prahovaní sa používa HSV farebný model (Obr. 5). HSV je skratka z anglických slov hue, saturation a value. Hue je prevládajúci farebný odtieň, saturation vyjadruje sýtosť farby a value obsahuje informáciu o množstve bieleho svetla. Pri prahovaní sa berú do úvahy iba hue a saturation, pretože sme chceli docieliť aby bolo prahovanie odolnejšie voči zmene osvetlenia. Výsledný vyprahovaný obraz sa uloží do nového objektu triedy Bitmap, z ktorého sa ďalej rozpoznávajú základné tvary ako kruh a obdĺžnik. Tento proces rozpoznávania tvarov je popísaný v nasledujúcej kapitole.



Obr. 5 HSV farebný model

Pre prahovanie sa používajú dva prahy. Jeden prah je pre loptu a druhý prah pre bránu. Prah je určený minimálnou a maximálnou hodnotou, pričom tieto hraničné hodnoty sme určili experimentálne. Pixely, ktorých hodnoty jasov spĺňajú jeden z dvoch prahov sú vykreslené príslušnou farbou daného prahu (Obr. 6). Všetky ostatné pixely sú vykreslené čiernou farbou.

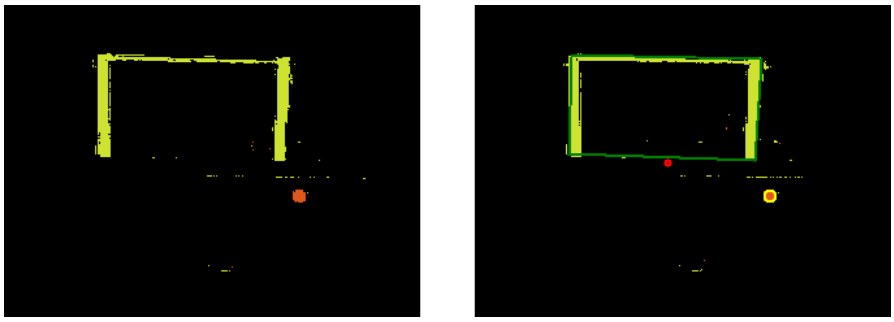


Obr. 6 Originál obraz vľavo a vyprahovaný obraz vpravo s rozlíšením 320x240 pixlov

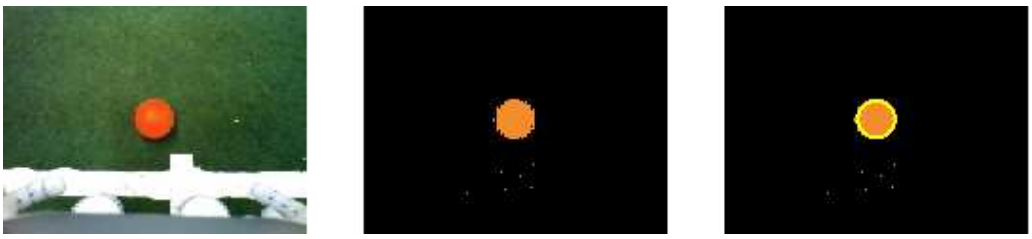
B. Rozpoznávanie tvarov

V riadiacom systéme vyhľadávania a orientácie robota Nao bolo potrebné vedieť súradnice lopty a stred brány. Keďže obrys brány je obdĺžnikového tvaru a obrys lopty je kruhového tvaru, stačilo rozpoznávať tieto dva typy tvarov.

Na rozpoznávanie tvarov sme použili framework AForge.NET, ktorý je open source [3]. Konkrétne sme využili knižnicu AForge.Imaging, ktorá obsahuje rozličné rutiny pre spracovávanie obrazu, ktoré uľahčujú prácu programátorovi. Funkcie pre rozpoznávanie tvarov z knižnice AForge.Imaging vyžadujú vyprahovaný obraz s čiernym pozadím (Obr. 6). Pri bráne bolo potrebné rozpoznať jej stred, aby robot vedel približnú polohu brány. Stred brány sa určí ako stred dlhšej strany obdĺžnika, ktorý opisuje bránu. Na obraze je tvar opisujúci bránu vykreslený zelenou farbou a stred brány je vykreslený červeným kruhom (Obr. 7). Lopta je označená žltým kruhom (Obr. 7, Obr. 8).



Obr. 7 Vyprahovaný obraz naľavo a rozpoznanie tvarov napravo s rozlíšením 320x240 pixlov



Obr. 8 Originál obraz naľavo, vyprahovaný obraz v strede a rozpoznanie kruhu napravo

Na Obr. 8 úplne vľavo je lopta tesne pred robotom. Pri prahovaní lopty, ktorá sa nachádzala tesne pred robotom sa často stávalo, že vplyvom zlého osvetlenia respektíve zmenou osvetlenia nebolo vidno celú loptu a preto bolo problematické rozpoznať kruh. Taktiež ak na obraze bola iba časť lopty, program ju nevedel rozpoznať ako kruh. Kvôli týmto problémom sme využili funkcie na rozpoznávanie neurčitých tvarov z knižnice AForge.Imaging. Pre tieto funkcie bolo dostačujúce ak na obraze bolo vyprahovaných aspoň niekoľko pixlov, z ktorých sa vytvoril objekt neurčitého tvaru (Obr. 11).



Obr. 9 Originál obraz naľavo, vyprahovaný obraz v strede a rozpoznanie neurčitého tvaru napravo

Pri lopte je potrebné poznať jej súradnice na obraze, aby sa robot vedel okolo nej otáčať. Pomocou funkcií na rozpoznávanie tvarov z framework-u AForge.NET, kde každý rozpoznaný tvar je chápaný ako objekt, sa dajú získať súradnice stredov týchto objektov. Týmto spôsobom sme získali súradnice lopty, ktoré sú vstupom do neurónovej siete FALCON.

IV. ROZHODOVACÍ BLOK

Hlavnou úlohou rozhodovacieho bloku je na základe dát z robota rozhodovať čo má robot spraviť. Z robota sa získavajú 2 typy dát:

- obraz (video)
- natočenia kĺbov

Z obrazu sa pomocou prahovania a rozpoznávania tvarov dá zistiť poloha brány a poloha lopty. Natočenia kĺbov sa dajú zistiť pomocou funkcie getAngles(object names, bool useSensors) z knižnice naoqidotnetvc90.dll, kde names sú názvy kĺbov a useSensors určuje či sa hodnoty kĺbov majú získať pomocou senzorov alebo nie.

Celá logika rozhodovacieho bloku je v programe implementovaná pomocou časovačov (timers). Na začiatku celej sekvencie, keď je robot mimo lopty a nevie kde sa lopta nachádza, sa robot snaží nájsť loptu otáčaním hlavy do strán, pričom je zapnutá spodná kamera. Dôvody, prečo používame v tejto časti spodnú kameru budú vysvetlené v kapitole V. Hlava robota sa horizontálne otáča približne od -57 stupňov do +57 stupňov. Hodnoty uhlov sme volili na základe experimentov, pričom ak by bol väčší rozsah otáčania, robot by nevidel loptu kvôli ramenám. Toto otáčanie zabezpečuje kĺb HeadYaw. Kĺb HeadPitch je nastavený na -9.6 stupňa. V prípade, že v danej oblasti robot nenájde loptu, celý robot sa otočí približne o 90 stupňov do pravej strany. Po nájdení lopty sa zistí uhol natočenia kĺbu HeadYaw a hlava sa nastaví na tento uhol tak, aby robot stále videl loptu cez spodnú kameru. Následne sa zapne časovač, ktorý ma na starosti sledovanie lopty chôdzou (kapitola V). Po tom ako robot úspešne príde k lopte na vzdialenosť asi 5 centimetrov, sledovanie lopty sa ukončí stopnutím príslušného časovača.

Následne robot začne hľadať bránu otáčaním hlavy, pričom je zapnutá vrchná kamera. Hlava robota sa zasa horizontálne otáča približne od -116 stupňov do +116 stupňov. Kĺb HeadPitch je teraz nastavený na nula stupňov. Robot nájde bránu vtedy, keď sa pomocou rozpoznávania tvarov vykreslí stred brány. Ak je na obraze viditeľný stred brány, získa sa uhol natočenia kĺbu HeadYaw a hlava robota sa natočí na získaný uhol. Tým získame uhol natočenia brány vzhľadom na robota. Na základe tohto uhla sa potom rozhodne, či sa robot začne otáčať do pravej strany alebo do ľavej strany. V prípade, že je uhol natočenia brány záporný, robot sa

začne otáčať do pravej strany. V opačnom prípade sa bude otáčať do ľavej strany. Na otáčanie robota okolo lopty sa používa neurónová sieť FALCON. Otáčanie robota okolo lopty je podrobne popísané v kapitole VI. Keď je robot správne natočený na bránu, skontroluje sa vzdialenosť od lopty. Ak je vzdialenosť väčšia ako päť centimetrov, opäť sa spustí tracker na sledovanie lopty chôdzou, s tým že robot sa pohybuje smerom dopredu po malých krokoch. Ak je vzdialenosť dostatočne malá, môže sa spustiť kop (kapitola VII).

V. SLEDOVANIE LOPTY CHÔDZOU

Pre sledovanie lopty chôdzou sme využili tracker zo spomínanej knižnice naoqidotnetvc90.dll. Nachádza sa v module s názvom RedBallTrackerProxy. Tracker je vo všeobecnosti zariadenie na sledovanie objektov pohybujúcich sa v trojrozmernom prostredí. Ako príklad môžeme uviesť kamerový tracker, solárny tracker, tracker letov komerčných liniek a podobne. Tento tracker je v podstate kamerový tracker, s tým že loptu sleduje otáčaním hlavy robota. Robot natáča hlavu za loptou tak, by bola lopta čo najbližšie k stredu obrazu. Lopta sa rozpoznáva len na základe farby, čo je v niektorých situáciách dosť nevýhodné. Tracker spustíme pomocou metódy startTracker() a zastavíme metódou stopTracker().

Spomínaný tracker sme sa rozhodli použiť, pretože je dostačujúci a aj v prípade, že by sme spravili vlastný tracker, nebol by o mnoho lepší ako tento. Hlavným dôvodom je kvalita obrazu získaná z robota. Ak je lopta od robota vzdialená viac ako dva až dva a pol metra, robot ju na rozlíšení 160x120 pixlov skoro vôbec nevidí. Rozdielom nie je ani rozlíšenie 320x240 pixlov, ktoré je taktiež nedostačujúce, kvôli malému počtu vyprahovaných pixlov lopty. Aj napriek tomu, že sa vyprahuje zopár pixlov, ktoré predstavujú loptu, na obraze je ich tak málo, že sú viditeľné ako šum. Vlastný tracker by bolo preto výhodné spraviť v prípade, že veľkosť rozlíšenia by bola minimálne 640x480 pixlov. Na tomto rozlíšení je lopta vykreslená väčším počtom pixlov, ktoré by už boli rozpoznané ako kruh (Obr. 8) alebo objekt neurčitého tvaru (Obr. 9). Aj keď obraz v tomto rozlíšení je možné získať z robota, bolo by to nevyhovujúce, pretože FPS je menej ako jeden čo je pre real-time aplikáciu neprípustné.

Spomínaný tracker z knižnice naoqidotnetvc90.dll sleduje loptu hlavou, pričom bolo potrebné spraviť tracker, ktorý by sledoval loptu chôdzou. Tracker na sledovanie lopty chôdzou sme spravili tak, že počas sledovania lopty hlavou sa zisťujú hodnoty uhlov kĺbu HeadYaw a na základe tohto uhla sa nastavuje parameter theta vo funkcii setWalkTargetVelocity(float x, float y, float theta, float frequency). Parameter theta určuje veľkosť uhla každého kroku robota. Funkcia setWalkTargetVelocity zabezpečuje chôdzu, pričom je to tzv. non-blocking volanie, to znamená, že v prípade ak sa zavolá daná funkcia s novými parametrami, robot ju začne hneď vykonávať. Existuje aj blocking volanie, ktoré funguje tak, že ak sa zavolá funkcia, ktorá má toto blocking volanie a práve sa vykonáva rovnaká funkcia, ktorá bola zavolaná pred ňou, zaradí sa táto funkcia do zásobníka a po ukončení práce prvej funkcie sa začne vykonávať nasledujúca funkcia v zásobníku. Vstupné parametre funkcie setWalkTargetVelocity sú:

- x – veľkosť kroku v predo zadnom smere [-1.0 , 1.0]
- y – veľkosť kroku v bočnom smere [-1.0 , 1.0]
- theta – veľkosť uhla natočenia kroku [-1.0 , 1.0]
- frequency – frekvencia krokov [0.0 , 1.0]

Uhol theta sa nastavuje v závislosti od veľkosti natočenia kĺbu HeadYaw, pričom veľkosť uhla theta sa vypočíta tak, že hodnota kĺbu HeadYaw sa predelí deväťdesiatimi stupňami. V prípade, že uhol kĺbu HeadYaw je väčší ako ±90 stupňov, theta je buď 1 alebo -1, v závislosti od znamienka. Výsledná hodnota je v požadovanom rozsahu [-1.0, 1.0], pričom ak je lopta viac na stranách, robot sa natáča rýchlejšie ako keď je lopta pred ním, čo je veľmi vyhovujúce. Parameter y je stále nulový a parameter x je stále kladný, čo znamená, že robot sa pohybuje výlučne dopredu. Tento tracker sa využíva v dvoch prípadoch:

- pri približovaní robota k lopte
- pri približovaní robota k lopte pri kope

V oboch prípadoch je zapnutá spodná kamera, kvôli zjednodušeniu problému. Zistili sme, že ak zapneme tracker z modulu RedBallTrackerProxy a prepneme spodnú kameru, sledovanie lopty funguje pomocou tejto kamery. Spodnú kameru využívame preto, lebo odpadá nutnosť prepínania kamier počas sledovania lopty chôdzou. Prepínanie kamier by bola nutnosť keby sa používala vrchná kamera, pretože pomocou nej nie je vidieť skoro jeden meter pred robotom, aj v prípade, že robot má natočenú hlavu čo najviac dole. Pri približovaní robota k lopte a tiež aj pri kope robot zastaví, ak je uhol kĺbu HeadPitch väčší ako 26.5 stupňa. Tento uhol zodpovedá približnej vzdialenosti päť centimetrov medzi loptou a robotom. Pri približovaní robota k lopte sa robot pohybuje pomalšie smerom dopredu, ak je robot približne pol metra od lopty, čomu zodpovedá uhol približne 15 stupňov. Parameter x je v tomto prípade nastavený na hodnotu 0.5, inak je nastavený na 1.

Dosť veľká nevýhoda trackera z modulu RedBallTrackerProxy je, že loptu rozpoznáva na základe farby. Problém nastáva, ak je na obraze viac objektov, ktoré majú podobnú farbu ako lopta. Tento problém sme odstránili rozpoznávaním tvarov. Ak je na obraze viac objektov, robot sa snaží primárne sledovať kruh. Ak je jeden objekt na obraze a je sfarbený ako lopta, robot ho bude sledovať, pretože sa predpokladá, že daný objekt je lopta. V tomto prípade nezáleží či to je kruh alebo objekt neurčitého tvaru, dôležité je, aby bol rozpoznaný ako jeden zo spomínaných tvarov. Kvôli týmto obmedzeniam sa robot občas zastaví, čo spomaľuje priebeh celej sekvencie, ale je to potrebné aby robot nezačal sledovať nesprávny objekt.

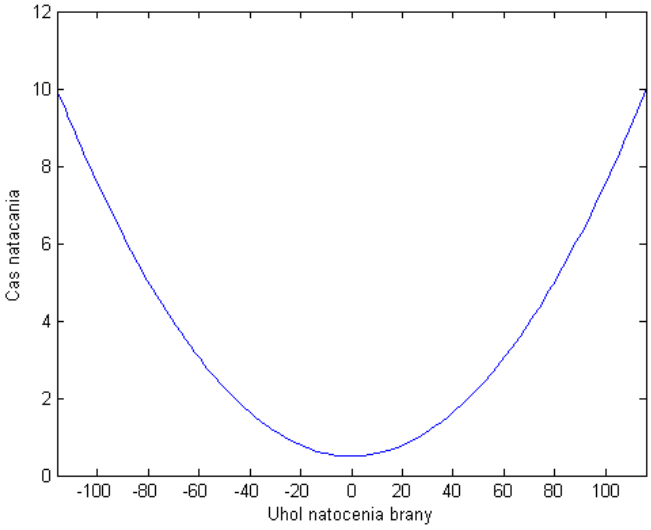
VI. OTÁČANIE ROBOTA OKOLO LOPTY

Ak je robot dostatočne blízko pri lopte, je potrebné aby sa správne natočil na bránu, pričom lopta aj brána by mali byť stále pred ním (Obr. 1). Toto natáčanie je dôležité kvôli presnosti kopu na bránu. Na natáčanie robota okolo lopty je použitá neurónová sieť FALCON.

Po tom ako robot úspešne príde k lopte na vzdialenosť približne päť centimetrov (kapitola V), zistí sa uhol natočenia brány otáčaním hlavy do strán. Hlava sa otáča v kĺbe HeadYaw v rozmedzí od -116 stupňov do +116 stupňov. Kĺb HeadPitch je nastavený na nula stupňov. Akonáhle sa na obraze vykreslí stred brány, zistí sa uhol kĺbu HeadYaw, pričom tento uhol je uhol natočenia brány vzhľadom na robota. Na základe veľkosti uhla natočenia brány sa určí, do ktorej strany a ako dlho sa bude robot otáčať okolo lopty.

Ak je uhol natočenia brány záporný, robot sa začne otáčať do pravej strany. V opačnom prípade sa bude otáčať do ľavej strany. Pre každý smer je použitá jedna neurónová sieť FALCON, ktorá je naučená len pre daný smer. Učenie siete prebiehalo off-line, to znamená, že najprv sa získali trénovacie dáta a potom sa sieť učila. Naučená sieť sa uložila a ďalej sa už pracovalo s naučenou sieťou. Tento proces sa vykonal pre oba smery a tým sme získali dve siete. Pri učení boli na vstupe do siete súradnice lopty na obraze a na výstupe boli parametre theta a y. Tieto dva parametre sú vstupnými parametrami pre funkciu setWalkTargetVelocity, ktorá už bola popísaná v kapitole V.

Následne je potrebné zistiť, ako dlho sa má robot pomocou neurónovej siete FALCON otáčať. V prípade, že by bol čas na otáčanie robota zle nastavený, mohlo by sa stať, že robot by sa nikdy správne nenatočil na bránu. Ak by sa nastavil konštantný čas na otáčanie robota okolo lopty, tento čas by musel byť dostatočne malý na to, aby sa jednak robot dostal do požadovanej pozície ale aj aby požadovanú pozíciu neprešiel. Týmto spôsobom by bola celá sekvencia natáčania veľmi zdĺhavá. Z úlohy vyplýva, že čím je v absolútnej hodnote uhol natočenia brány väčší, tým je aj čas na otáčanie robota dlhší. Naopak, ak je uhol natočenia brány malý, aj čas na otáčanie robota okolo lopty je krátky. Preto sme na odhad dĺžky času, keď sa robot otáča okolo lopty, použili kvadratickú funkciu (Obr. 10), ktorú sme navrhli experimentálne.



Obr. 10 Kvadratická funkcia zobrazujúca čas natáčania robota okolo lopty

Na x-ovej osi je uhol natočenia brány v spomínanom rozmedzí od -116 do +116 stupňov. Na y-ovej osi je čas natáčania robota okolo lopty. Počas tohto času sa získava výstup z naučenej siete a riadi sa pohyb robota okolo lopty. Po uplynutí času robot zastane a opäť začne hľadať bránu

otáčaním hlavy. Celý cyklus sa opakuje pokiaľ uhol natočenia brány nie je menší ako ±10 stupňov. Keďže ihrisko je dosť malé, tento uhol je dostačujúci.

A. Učenie neurónovej siete FALCON

Počas učenia neurónovej siete FALCON sme zistili, že pri úlohe natáčania robota okolo lopty je najlepšie ak sú tri funkcie príslušností na oboch vstupoch a päť funkcií príslušností na oboch výstupoch. V ostatných prípadoch sa sieť stále naučila horšie. Preto sme pracovali iba s týmito počtami funkcií príslušnosti. V programe sme ohraničili maximálny počet funkcií príslušností na desať a minimálny počet na tri. Väčší a ani menší počet funkcií príslušností podľa [4] nemá zmysel.

B. Tvorba trénovacej množiny

Pri učení neurónovej siete FALCON sa používa algoritmus spätného šírenia chyby. Tento algoritmus sa využíva pri parametrickom učení. Patrí do skupiny kontrolovaného učenia, to znamená, že ku každému vstupu trénovacej množiny je pridelený požadovaný výstup. Aby sa neurónová sieť FALCON bola schopná naučiť natáčať robota okolo lopty, bolo potrebné vytvoriť trénovacie dáta, na ktorých by sa sieť učila.

Trénovacie dáta sme vytvorili tak, že pomocou joystick-u (Obr. 11) sme natáčali robota okolo lopty niekoľko desiatok sekúnd. Robot mal nastavenú hlavu na konštantný uhol, ktorý sa počas získavania dát nemenil. Takisto aj vo fáze života neurónovej siete FALCON bola hlava robota nastavená na rovnaký uhol ako vo fáze učenia. KĺbHeadYaw bol nastavený na nula stupňov a kĺb HeadPitch približne na 18 stupňov.



Obr. 11 Joystick na tvorbu trénovacích dát

Počas ovládania robota joystick-om sa ukladajú súradnice lopty a parametre y a theta z funkcie setWalkTargetVelocity, ktorá bola podrobne popísaná v kapitole V. Súradnice lopty sú vstupy a parametre y a theta sú výstupy. Nahrávanie trénovacej množiny sa jednoducho spustí stlačením a držaním hlavného tlačidla na joystick-u (Obr. 11). Akonáhle sa tlačidlo uvoľní, nahrávanie je zastavené, ale robot je aj naďalej ovládaný joystick-om. Pri ďalšom stlačení a držaní tlačidla sa pokračuje v ukladaní dát bez straty predošlých dát. Takto môžeme pokračovať až kým v programe nestlačíme v programe tlačidlo Save training set, ktoré nám uloží celú trénovaciu množinu. Ak opäť stlačíme tlačidlo na joystick-u, začne sa nahrávať nová trénovacia množina a celý cyklus sa opakuje. Tento spôsob nahrávania trénovacej množiny sa veľmi dobre osvedčil.

VII. KOP

Ideálna vzdialenosť medzi nohou robota a loptou je pri kope približne päť centimetrov. V prípade, že je vzdialenosť väčšia, robot sa pomalým kráčaním priblíži k lopte na požadovanú vzdialenosť pomocou trackera (kapitola V). Kopnutie do lopty je statická sekvencia pohybov, to znamená, že jednotlivé natočenia kĺbov sú dopredu nastavené a nijako sa nemenia. Ak je robot správne natočený na loptu a na bránu, lopta skončí v bráne. V tejto práci je spravený kop ľavou a pravou nohou, pričom závisí od polohy lopty, ktorý kop sa spustí. Pri kope pravou nohou sa používajú kĺby RAnkleRoll, LAnkleRoll, RHipRoll, RHipPitch. Pri kope ľavou nohou sa používajú kĺby LAnkleRoll, RAnkleRoll, LHipRoll, LHipPitch.

Určenie polohy lopty pri kope sa realizuje pomocou súradníc lopty. V strede obrazu sú súradnice [0,0]. Ak je lopta viac na pravej strane obrazu, x-ová súradnica je kladná a to znamená, že lopta je bližšie pri pravej nohe. V opačnom prípade je bližšie pri ľavej nohe. Ak sa lopta nenachádza presne pred nohou, robot sa malými úkrokmi do strán približuje k lopte až do chvíle, kým sa lopta nenachádza presne pred nohou. Od pozície lopty záleží, do ktorej strany začne robot robiť úkroky. Ak je lopta bližšie pri pravej nohe, robot bude robiť úkroky tak, aby bola lopta čo najpresnejšie pred pravou nohou. V opačnom prípade zas bude robiť úkroky tak, aby bola lopta čo najpresnejšie pred ľavou nohou. Ak je robot správne nastavený, spustí sa sekvencia pohybov realizujúca kop.

VIII. Obmedzenia

V tomto navrhnutom a implementovanom riadiacom systéme vyhľadávania a orientácie robota Nao je zopár obmedzení, ktoré zväčša súvisia s hardvérom. V nasledujúcej časti sú popísané všetky tieto obmedzenia.

Asi najväčšie obmedzenie je, že počas hľadania lopty môže byť lopta od robota vzdialená maximálne dva a pol metra. Pri väčšej vzdialenosti ju robot jednoducho nevidí, respektíve pri spracovaní obrazu je počet pixlov, ktoré predstavujú loptu veľmi malý (kapitola V).

Ďalšie obmedzenie súvisí so vzdialenosťou robota od brány. Ak je robot veľmi blízko brány, nedokáže ju rozpoznať, pretože vidí iba jej časť. Preto musí byť robot pri kope vzdialený od brány aspoň jeden meter.

Keďže robot je ovládaný počítačom cez Wi-Fi, je tam malé oneskorenie, ktoré v niektorých situáciách môže viesť k určitým nepresnostiam. Napríklad pri hľadaní brány keď robot otáča hlavu do strán sa môže stať, že robot posunie hlavu trocha ďalej a uhol natočenia brány nie je taký presný. Tieto obmedzenia hlavne súvisia s bezdrôtovým pripojením a dajú sa odstrániť napríklad pripojením robota na sieťový kábel alebo výmenou Wi-Fi antény v robotovi.

Posledným obmedzením je, že ak sa počas vykonávania celej sekvencie niečo stane, napríklad dôsledkom oneskorenia Wi-Fi si robot trocha odkopne loptu alebo robot spadne na zem, je potrebné spustiť celú sekvenciu ešte raz.

IX. Experimenty

Cieľom týchto experimentov bolo otestovanie funkčnosti navrhnutého riadiaceho systému vyhľadávania a orientácie robota na reálnom ihrisku. V experimentoch boli použité dve neurónové siete FALCON, každá pre iný smer otáčania. Tieto siete boli pred experimentmi naučené a počas experimentu sa nijako nemenili.

Na začiatku každého experimentu boli robot aj lopta umiestení na inom mieste na ihrisku. Počas experimentu sa meral čas, za ktorý prejde robot z počiatočnej pozície k lopte (tracker V) a čas otáčania robota Nao okolo lopty až do chvíle, kým robot nebude nastavený priamočiaro na loptu a bránu (kapitola VI). Zároveň sa zaznamenáva či robot strelí gól alebo nie. Treba poznamenať, že čas otáčania robota okolo lopty dosť závisí na tom, ako rýchlo dokáže robot nájsť bránu. V tabuľke v nasledujúcej časti sú popísané časy trackera a otáčania robota okolo lopty. Na nasledujúcom obrázku je fotka jedného z experimentov s počiatočnými pozíciami robota a lopty na ihrisku.



Obr. 12 Počiatočná pozícia robota a lopty na ihrisku experimentu č. 1.

Tab. 1
Výsledky experimentov

Experiment č.	Čas trackera	Čas rotácie	Gól
1	25.5	16.4	áno
2	22	13.1	áno
3	22.4	23.6	áno
4	56	32	áno
5	22	11.1	nie
6	40.3	18	nie
7	19.9	12	nie
8	24.4	47.8	áno
9	23.3	15.3	áno
10	25	14	nie
11	21.6	19.5	áno
12	31.4	42.8	Nie
13	15.3	20.1	nie
14	24.4	47.8	áno
15	23	31.4	áno
16	25.7	45.3	áno
17	21	10	áno
18	29.5	42.6	nie
19	19.3	37.8	nie
20	26.4	34.1	nie

X. ZÁVER

V tejto práci je navrhnutý a naprogramovaný riadiaci systém vyhľadávania a orientácie robota Nao pre potreby kopu spolu s neurónovou sieťou FALCON. V experimentoch sú použité dve neurónové siete FALCON pre dva smery otáčania robota okolo lopty. Siete boli naučené pred experimentmi a počas experimentov sa už nemenili.

V práci sme použili 20 náhodných experimentov. Priemerný čas, za ktorý sa robot dostal z počiatočnej pozície k lopte je 25.92 sekúnd. Priemerný čas otáčania robota okolo lopty je 26.735 sekúnd. Úspešnosť strelenia gólu je 55 percent. Čas otáčania robota okolo lopty veľmi závisí od toho, ako rýchlo dokáže robot nájsť loptu, pretože pri otáčaní si robot kontroluje polohu brány. Ak sa náhodou začne hlava otáčať do opačnej strany ako leží brána, čas sa predlžuje. Smer otáčania hlavy robota sa volí náhodne. Strelenie gólu závisí od toho, či je robot správne natočený na loptu a bránu. Pri dolaďovaní svojej polohy sa robot približuje k lopte tak, aby bola lopta buď pred ľavou alebo pravou nohou. Práve pri tomto procese sa môže stať, že robot zaznamená, že lopta je na požadovanej pozícii aj keď v skutočnosti nie je. Tieto mylné hodnoty vznikajú dôsledkom nakláňania robota pri chôdzi. Dalo by sa to vyriešiť zastavením robota pokiaľ sa získajú údaje o polohe lopty ale to by celú sekvenciu oveľa viac spomalilo. Kvôli rýchlosti sme sa preto rozhodli, že robot sa bude pohybovať počas celého procesu získavania údajov o polohe lopty.

Myslím si, že výsledky experimentov sú v celku dobré, aj keď existuje veľa možných vylepšení, ktoré by tento systém urýchlili. Jedným z hlavných vylepšení by mohlo byť predspracovanie obrazu priamo v robotovi. Do počítača by sa potom posielali iba dôležité

informácie z obrazu, čo by značne zvýšilo FPS. Ďalšie vylepšenia by sa mohli týkať pamätania si pozície brány, lopty, robota respektíve aj ďalších hráčov.

POĎAKOVANIE

Táto práca bola vytvorená realizáciou projektu Rozvoj Centra informačných a komunikačných technológií pre znalostné systémy (kód ITMS projektu: 26220120030) na základe podpory operačného programu Výskum a vývoj financovaného z Európskeho fondu regionálneho rozvoja.

LITERATÚRA

[1] Lin, Ch. T. – Lee, G.: Neural Fuzzy Systems – A Neuro-Fuzzy Synergism to Intelligent Systems; Prentice Hall PTR, New York, 1996, s. 533-544, ISBN 0-13-235169-2
[2] Tomori, Z.: Počítačové videnie – dočasné učebné texty pre poslucháčov FEI TUKE; Košice, 2008
[3] AForge.NET framework; dostupné na internete: <http://www.aforgenet.com/ framework>
[4] Vaščák, J.: Fuzzy logika v regulácii – Podrobný úvod do problematiky fuzzy riadenia; 2008
[5] Nao Software Documentation; dostupné na internete: <http://www.aldebaran-robotics.com/documentation>

Optimalizácia procesov v stavebnej firme

Lenka ILAVSKÁ, František BABIČ

Katedra kybernetiky a umelej inteligencie, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

lenka.ilavska@student.tuke.sk, frantisek.babic@tuke.sk

***Abstrakt* — Optimalizácia podnikových procesov predstavuje dôležitý krok v celkovému zlepšeniu fungovania príslušného podniku alebo firmy. Na tento účel je možné použiť viacero prístupov ako napr. postupné zmeny v procesoch alebo okamžitá zmena od základov. V tomto článku bude popísaný postup práve pomocou druhej metódy s názvom BPR (Business Process Reengineering). Pomocou tejto metódy budú identifikované neefektívne procesy vo vybranej stavebnej firme, ktoré budú následne namodelované pomocou štandardu BPMN (Business Process Modeling Notation). Vytvorené procesné modely budú podrobené simuláciám a následným analýzam s cieľom, identifikovať potenciálne možnosti a miesta zlepšenia. V druhom kroku bude analyzovaná celková situácia vo firme pomocou metódy Balance Scorecard. Dosiahnuté výsledky predstavujú zaujímavú alternatívu ako zlepšiť a zefektívniť nielen hlavný proces výstavby bytových jednotiek, ale aj rôzne podporné a organizačné procesy vo firme.**

***Kľúčové slová* — Procesné modelovanie, Simulácie, Balance ScoreCard**

I. ÚVOD

A. Analýza súčasného stavu

Funkčné riadenie podnikov, ktoré rieši predovšetkým deľbu práce a je zamerané na dôsledky vzniku nedostatkov, sa pomaly dostáva do úzadia a nastupuje procesné riadenie, kde sú činnosti podniku usporiadané podľa nadväznosti v procesoch. Toto riadenie sa zameriava na zistenie príčin vzniku nedostatkov, a na základe týchto zistení je možné zmeniť neefektívne podnikové procesy na efektívne. Základom procesného riadenia je tvrdenie, že príčinou problémov v podniku sú práve prebiehajúce procesy a chyby v nich, ktoré je nutné riešiť a opraviť. Procesného riadenie je založené na zjednotení činností podniku a procesu ako celku. [1]

Pred samotným meraním a hodnotením procesov v podniku by sa mala uskutočniť práve ich simulácia. Tento krok tradične nasleduje po identifikácii príslušných procesov a vytvorení procesných modelov. Simulácia patrí k moderným spôsobom analýzy podnikových procesov, kde cieľom je reálne napodobniť ich chod, a následne preskúmať jeho správanie. Taktiež umožňuje testovať alternatívne verzie procesov a sledovať kritickú cestu za účelom porozumenia správania systému. Jej cieľom je poskytnúť údaje o slabších miestach v procesoch a spôsobe využitia zdrojov. [2]

Optimalizácia predstavuje zosúladenie zdrojov systému, aby sledované kritérium nadobúdalo extrémne hodnoty (minimum, maximum). Cieľom simulácie a optimalizácie je identifikovať zmeny, ktoré je potrebné urobiť v štruktúre skúmaných procesov. [2]

Všeobecný cieľ každého podniku je byť konkurencieschopný a v čo najdlhšom časovom horizonte. Plnenie tohto cieľa je možné sledovať pomocou metódy Balance Scorecard[1], ktorá vznikla koncom 20. storočia v USA. Predstavuje komplexný prístup k podnikovému riadeniu na základe systému vyvážených ukazovateľov. Výhodou sú úzke vzťahy medzi strategickou a prevádzkovou úrovňou riadenia v týchto častiach firmy: financie, zákazníci, interné procesy, učenie a rast. [3]

Článok je rozdelený do troch hlavných častí, kde prvá popisuje identifikovaný problém a možné spôsoby riešenia; druhá časť je venovaná konkrétnym krokom pomocou vhodných metód procesného modelovania a simulácií, takisto techniky Balance Scorecard; posledná sumarizuje dosiahnuté výsledky a potenciál ich nasadenia do praxe.

[1] http://www.balancedscorecard.org

B. Identifikácia riešeného problému

Tento článok popisuje spôsob riešenia problému efektívnej realizácie zákazok v stavebnej firme prostredníctvom optimálneho využívania materiálnych a ľudských zdrojov. Navrhované riešenie je zamerané na poznanie vnútorných procesov, ktoré prispievajú k dosiahnutiu stanovených cieľov firmy, čiže ide nielen o výrobné (stavebné) procesy, ale aj organizačné alebo podporné.

Identifikovaný problém bude riešený pomocou vybraných vhodných softvérových prostriedkov na základe dát popisujúcich proces výstavby 2x8 bytových jednotiek. Prostredníctvom týchto dát budú vytvorené príslušné procesné modely, navrhnuté a spustené simulácie a analyzované dosiahnuté výsledky z pohľadu efektívnejšieho manažmentu zdrojov a času. Cieľovú skupinu predstavujú riadiace orgány podniku ako riaditeľ podniku a členovia predstavenstva, ktorých úlohou je určovať celkovú stratégiu firmy, ciele a spôsoby ich dosiahnutia. Ďalej môžu dosiahnuté výsledky využívať aj vedúci pracovníci jednotlivých oddelení, ktorých úlohou je prenášať stanovenú stratégiu firmy do konkrétnych podnikových procesov.

C. Nástroje

QPR ProcessGuide[2] reprezentuje interaktívny nástroj, ktorý umožňuje rýchlo a ľahko modelovať, monitorovať a analyzovať podnikové procesy. Umožňuje taktiež dokumentovať a simulovať podnikové procesy v podniku, vďaka čomu ich vieme neustále zlepšovať.

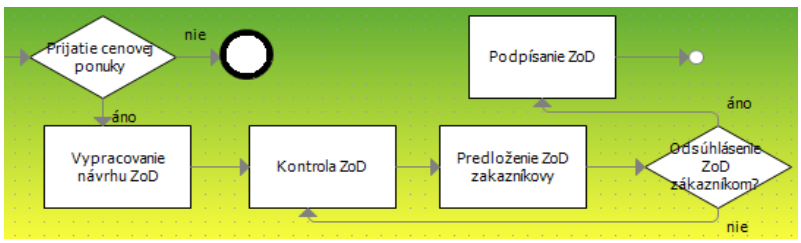
QPR ScoreCard[3] predstavuje pre kompletné riešenie procesov a riadenie výkonu pomocou metódy Balance ScoreCard. V tomto programe máme aktuálny prehľad o stave podniku a to pomocou reportov a grafov. Takisto umožňuje rozpracovať strategické ciele podniku do akčných plánov.

II. RIEŠENIE

A. Procesný model

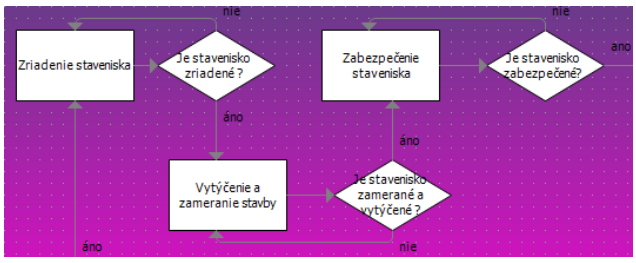
Prvým krokom bolo vytvorenie procesného modelu skúmaných procesov na základe informácií získaných z firmy. Tento model popisuje činnosť troch organizačných jednotiek firmy, t.j. Obchodno-technický úsek, Nákup a Výrobný úsek. Výsledný model obsahuje 12 podprocesov, z ktorých si ako ukážku predstavíme dva.

Podpísanie dokumentu „Zmluva o dielo (ZoD)“ predstavuje dôležitý krok v procese prípravy stavebnej činnosti, viď. Obr.1.



Obr. 1 Časť procesu popisujúca podpísanie Zmluvy o dielo.

V druhom prípade ide o prípravu staveniska, aby sa mohlo začať s konkrétnou výstavbou (Obr.2). Stavbyvedúci je povinný pri realizácií prác zabezpečiť výkopy tak, aby nedošlo k ohrozenie života a zdravia a taktiež zabezpečiť stavenisko proti vniknutiu cudzích osôb.

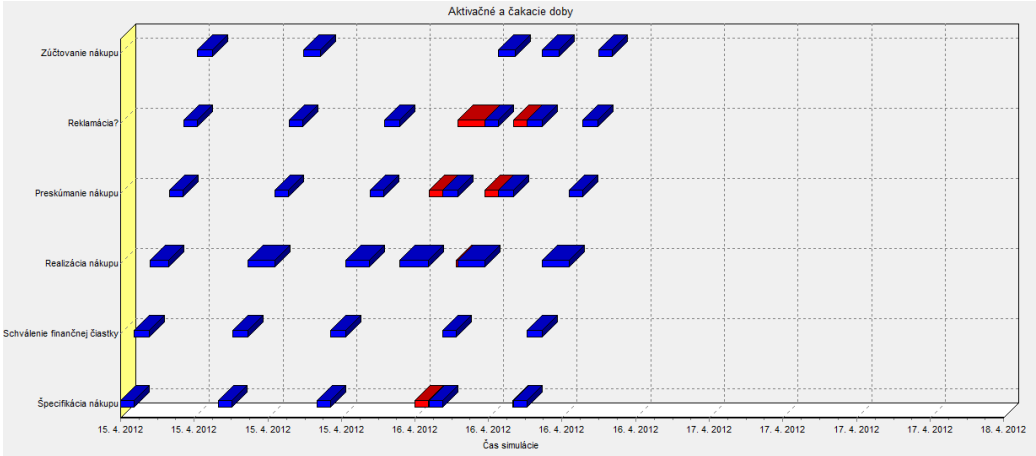


Obr. 2 Časť procesu popisujúca prípravu staveniska.

[2] http://www.qpr.com/products/qpr-processdesigner.htm
[3] http://www.qpr.com/solutions/balanced-scorecard-solution.htm

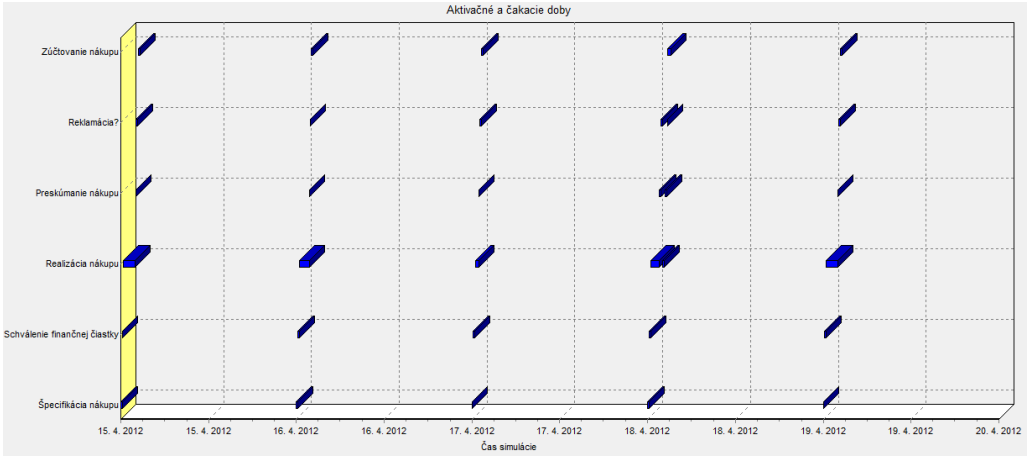
B. Simulácie

Každá simulácia je ako samostatný projekt. Je nutné definovať úlohy, navrhnúť ich nadväznosť, určiť potrebné zdroje (ľudské a výpočtové) a financie. Výsledkom je prehľad ako sú využívané zdroje v čase a aké sú náklady. Pre vyššiu prehľadnosť a orientáciu vo výsledkoch sa simulácie vykonávajú na častiach procesného modelu. V nástroji QPR ProcessGuide je pre tieto účely nutné zadať parametre ako napr. *doba spracovania*, ktorá slúži na meranie času medzi začiatkom a koncom procesu; *čas trvania* na meranie efektívnosti procesu; a *náklady* na vyčíslenie všeobecných nákladov spojených s procesom. Každý krok v procese má určenú práve jednu číselnú hodnotu nákladu.



Obr. 3 Výsledky simulácie podprocesu „drobné nákupy“ - pôvodná verzia.

Prvým krokom pri tvorbe simulácií je určiť, kto je vlastník procesu a aké merania budú v danej simulácií použité. V prípade podprocesu „drobné nákupy“ je vlastníkom stavbyvedúci a ako merania boli zvolené doba spracovania, čas trvania a náklady. Vlastnosti simulácie obsahujú dobu spracovania, frekvenciu simulácie, vstupné a výstupné podmienky. V prípade špecifikácie nákupu je doba spracovania približne 10 minút, frekvencia aktivácie je každý deň. Počet aktivácií je 5, čo znamená, že prebehne 5 simulácií. Výsledkom tejto simulácie sú odhalené kritické miesta podprocesu (Obr. 3). V tomto prípade išlo o nedostupnosť potrebných zdrojov v danom čase (červená farba), keďže stavbyvedúci bol veľmi vyťažený a nestíhal plniť dané úlohy. Zlepšenie v tomto prípade spočíva v presune menej dôležitých činností na asistenta stavbyvedúceho. Výsledok je zobrazený na nasledujúcom obrázku.



Obr. 4 Výsledky simulácie podprocesu „drobné nákupy“ - upravená verzia.

Ďalším príkladom je podproces Stavenisko, pri ktorom samotná simulácia neodhalila žiadne kritické miesta. Viaceré kroky tohto podprocesu však majú rovnaké vstupné zdroje, čiže ich zlúčenie do jedného kroku prinieslo skrátenie času realizácie, viď. Tab.1.

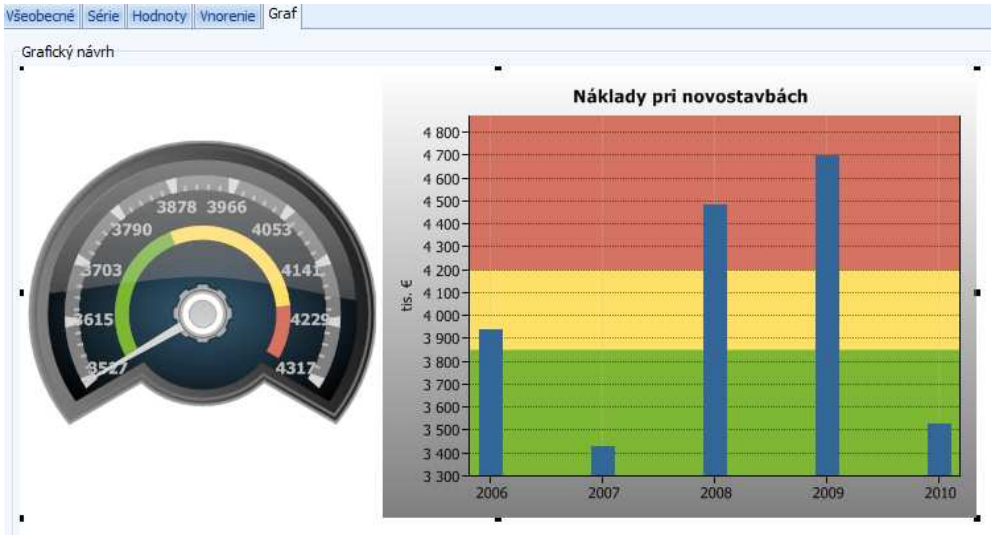
Tabuľka 1
Stavenisko pred/po

Krok	Pred		Po	
	Trvanie hod/min	Náklady %	Trvanie hod/min	Náklady %
Preberanie	15/05	28,79	15/05	29,45
Pripravené?	01/15	02,22	01/15	02,34
Zriadenie	17/13	32,77	17/13	33,50
Zriadené?	01/10	02,11	01/10	02,25
Vytýčenie	**04/14**	**07,92**		
Vytýčené?	**01/06**	**02,02**	**08/25**	**16,19**
Zameranie	**03/56**	**06,82**	**00/43**	**00,94**
Zamerané?	**01/03**	**02,48**		
Zabezpečenie	06/46	12,36	06/46	12,70
Zabezpečené?	01/31	02,51	01/31	02,63
SPOLU	**52/26**	100	**51/28**	100

C. QPR Scorecard

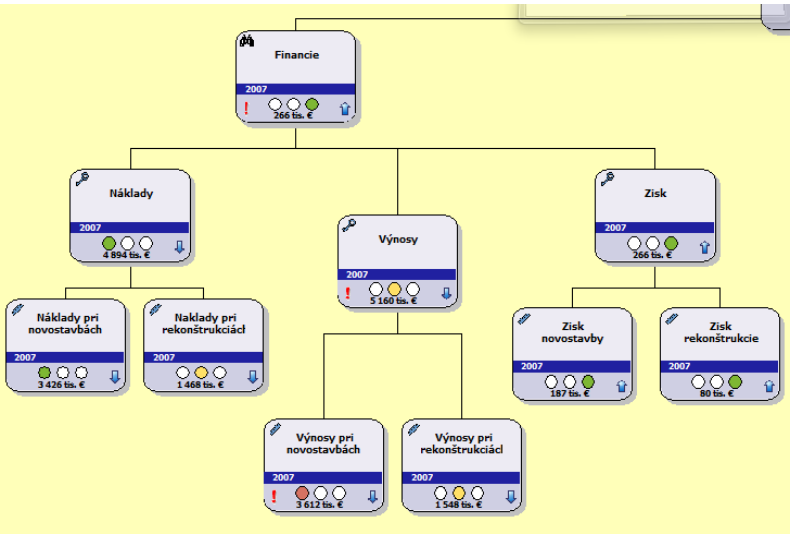
Prvým krokom pri tvorbe prehľadného popisu stavu podniku je určenie podnikových ukazovateľov, ktoré určujú jeho chod. V našom prípade sme ako základné ukazovatele určili financie, motivácia zamestnancov, spokojnosť zákazníkov a vnútorné procesy. Pri financiách sa berie do úvahy zisk podniku, ktorý sa určuje rozdielom výnosov a nákladov. Pri motivácii zamestnancov je rozhodujúca ich spokojnosť, na ktorú vplýva hlavným podielom rast miezd. Spokojnosť zákazníkom je založená na ich spokojnosti s ponúkanými službami, počtu sťažností a počte samotných zákazníkov. Pre vnútorné procesy je dôležité nenavýšenie rozpočtu a dokončenie zákazky načas.

Druhým krokom tejto analýzy je vyplnenie údajov a určenie kritických a dobrých výsledkov. Výsledky sú graficky prezentované používateľovi (viď. Obr.5), ktorý je schopný zhodnotiť aktuálne fungovanie firmy (podniku), stanoviť ciele a stratégiu do budúcnosti, navrhnúť opatrenia na zlepšenia, atď. Ak je stĺpec v danom roku v zelenej časti, podnik dosiahol výborné výsledky; ak je v červenej časti, tak nesplnil svoj cieľ a nedosiahol dobré výsledky.



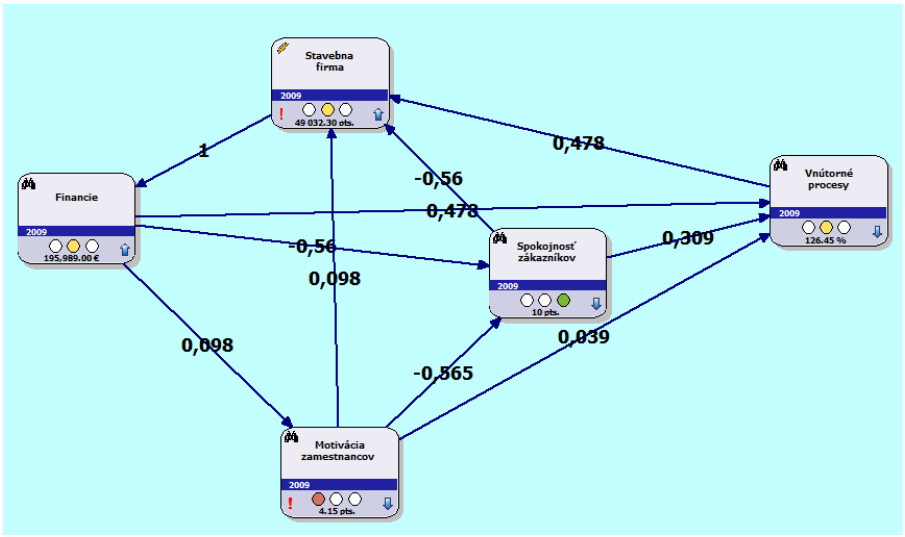
Obr. 5 QPR SC – graf zobrazujúci vynaložené náklady na novostavby v rokoch 2006 – 2010.

Výsledok je možné zobraziť aj pomocou prehľadnej stromovej štruktúry (viď. Obr.6), ktorá zobrazuje jednotlivé položky financií vo firme.



Obr. 6 QPR SC – graf zobrazujúc i štruktúru financií vo firme za rok 2007.

Softvérové prostredie QPR Scorecard umožňuje takisto návrh strategickej mapy, ktorá zobrazuje korelácie (od -1 po 1) medzi zvolenými ukazovateľmi SC. Strategická mapa podniku na Obr. 7 popisuje vzťahy medzi podnikom (stavebná firmy), financiami, motiváciou zamestnancov, spokojnosťou zákazníkov a vnútornými procesmi v roku 2009.



Obr. 7 Strategická mapa

III. ZÁVER

Tento článok popisuje spôsob optimalizácie procesov vo vybranej stavebnej firme z pohľadu časovej náročnosti a vynaložených nákladov. Pomocou vhodne navrhnutých simulácii boli v týchto procesoch identifikované kritické miesta a navrhnuté zlepšenia na ich odstránenie. Nové verzie modelov boli opätovné simulované s cieľom potvrdiť efektívnosť navrhnutých zlepšení. Pomocou metódy Balance Scorecard bol vytvorený prehľad o fungovaní firmy z pohľadu štyroch základných ukazovateľov, t.j. financie, motivácia zamestnancov, spokojnosť zákazníkov a vnútorné procesy. Prezentované výsledky sú len časťou analýzy, ktorá umožnila skúmanej firme optimalizovať svoje chod, znížiť investované náklady a zlepšiť ponúkané podmienky pre klientov.

POĎAKOVANIE

Tento príspevok vznikol za podpory KEGA projektu č. 065TUKE-4/2011 (50%) a realizáciou projektu Rozvoj Centra informačných a komunikačných technológií pre znalostné systémy (kód ITMS projektu: 26220120030) na základe podpory operačného programu Výskum a vývoj financovaného z Európskeho fondu regionálneho rozvoja (50%).

LITERATÚRA

[1] F. Kovář, H. Kožíšková, K. Hrazdilová, *Teorie průmyslových podnikatelských systémů II*. Zlín: Univerzita Tomáša Baťu v Zlíne, 2004

[2] J. Zavacký, *Procesný manažment v praxi manažéra*. SP Synergia, 2004.

[3] I. Hudáková, *Balance ScoreCard - cesta k úspechu v strategickom riadení podnikov*. Dostupné na internete: <www.pp.sk/WebDAU/main/ArchivArticleShow.aspx?ArtID=28723&print=1>.

Návrh a realizácia riadenia flexibilného montážneho podniku PLC automatom

Igor JADLOVSKÝ, Matej ČOPÍK, Ján SARNOVSKÝ

Katedra kybernetiky a umelej inteligencie, Fakulta elektrotechniky a informatiky,
Technická univerzita v Košiciach, Slovenská republika

i.jadlovsky@gmail.com, matej.copik@tuke.sk, jan.sarnovsky@tuke.sk

***Abstrakt* — Tento článok sa venuje návrhu a realizácii riadenia Flexibilného montážneho podniku na technologickej úrovni pomocou PLC automatu. Tento výrobný systém je súčasťou distribuovaného systému riadenia budovaného na KKUI. Pre komunikáciu s externými systémami sú vytvorené komunikačné kanály. Riadiace algoritmy sú budované k dosiahnutiu efektivity a spoľahlivosti výrobného systému. Pre riadenie opakovaných sekvencií sú navrhnuté a vytvorené podprogramy. V práci sa kladie dôraz na paralelné procesy výroby a ich synchronizáciu.**

***Kľúčové slová* — Flexibilný montážny podnik, distribuovaný systém riadenia, Allen Bradley, Rockwell Automation, PLC, frekvenčný menič, krokový motor, manipulátor, dopravníkový pás, sklad**

I. ÚVOD

Oblasť priemyselnej automatizácie je odvetvím kybernetiky, ktoré je pre dnešnú výrobnú sféru nenahraditeľné. Automatizované systémy dnes vykazujú vysokú spoľahlivosť a presnosť výroby pri veľmi nízkych časových a znížených finančných nárokoch. Takýmto výrobným systémom je aj model flexibilného montážneho podniku (FMP), umiestnený v laboratóriu katedry Kybernetiky a Umelej inteligencie. Riadenie technologickej úrovne distribuovaného systému je priamym riadením hardwarových členov. Preto sa budem zaoberať ich analýzou a princípmi ich riadenia.

II. POPIS FMP

A. Produkt FMP

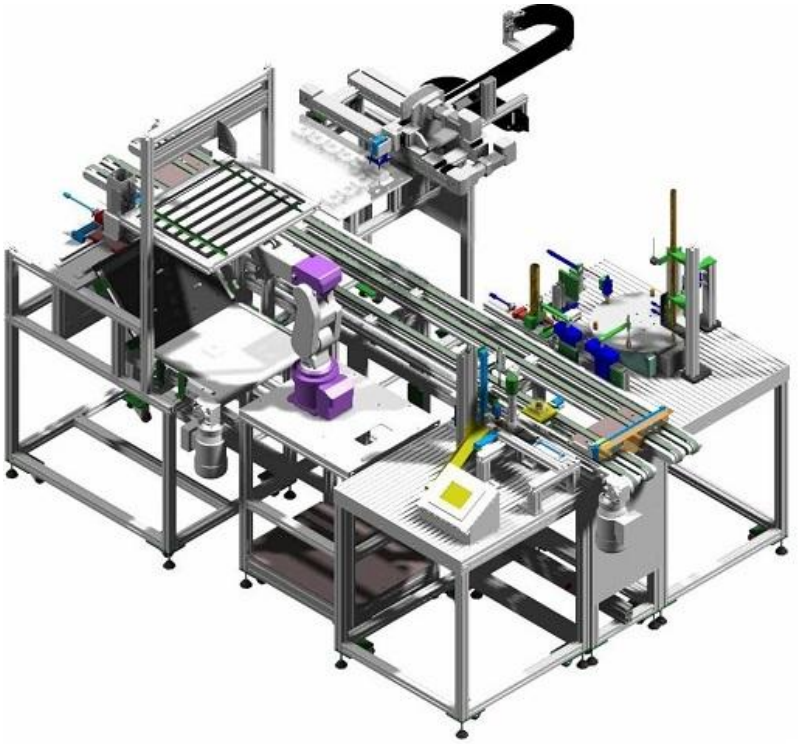
Produktom Flexibilného montážneho podniku je špecificky zapúzdrené ložisko. Ložisko je osadené do hliníkovej podstavy štvorcového tvaru s presne vyfrézovaným otvorom. Do stredového otvoru ložiska je vsunutá nylonová hriadeľka. Na hriadeľke je z vrchu osadený klobúčik. Produkt je vyrábaný v rôznych vyhotoveniach. Na výber sú hliníkové a železné klobúčiky, zelené a modré hriadeľky a ložiská dvoch rôznych rozmerov.

B. Funkcionalita FMP

Flexibilný montážny podnik je postavený na princípoch sériovej výroby. Na hlavných dopravníkových pásoch sú presúvané palety. Na paletách sú postupne jednotlivými postami vyskladávané produkty. Za výrobné posty sú považované post 1, post 3 a post 4. Post 0, post 2 a post 5 slúžia na manipuláciu s výrobkami v rámci linky a skladu. Každý výrobný post obsahuje materiálne vstupy organizované v zásobníkoch.

Výroba začína postom 1. Na poste 1 je do výrobného systému vkladaná podstava zo zásobníku. Post 2 pomocou robota presunie podstavu na paletu, ktorá je pristavená na hlavnom dopravníkovom páse označenom ako post 0. Paleta s podstavou sa presunie po páse k postu 3. Vstupný zásobník postu 3 obsahuje ložiská. Ložiská sú po otestovaní rozmerov osadené manipulátorom do podstavy na poste 0. Po dokončení operácie sa paleta presúva k postu 4. Na poste 4 sú do systému vložené a následne testované hriadeľky a klobúčiky. Pomocou manipulátora medzi postom 4 a 0 je hriadeľka osadená do ložiska a následne prikrytá klobúčikom. Tým je výrobná časť procesu ukončená. Hotový výrobok sa následne presunie k

postu 5. Tento post obsahuje výstupný sklad. Manipulátor uloží výrobok do výstupného skladu. Paleta pri tom ostáva na páse a je presunutá po hlavnom dopravníkovom páse na začiatok výroby.



Obr. 1 Flexibilný montážny podnik

Post 0

Slúži na presun paliet v rámci Flexibilného montážneho podniku. Počas výroby beží dopravníkový pás nepretržite. Palety sú pri prechode okolo postov zastavené stopermi a kotvami po dobu vkladania súčiastky daným postom. Následne je paleta uvolnená a presúvaná k nasledujúcemu postu. Presun paliet tvorí uzavretý cyklus. Po odobratí hotového výrobku sa paleta dostáva na začiatok výroby.

Cyklus postu 0 začína na prvom stanovisku. Tu sú palety skladované za sebou. Čakajú na uvolnenie nasledovného stanoviska. Po jeho uvolnení je jedna vždy jedna paleta pustená. Druhé stanovisko je pri poste 2. Paleta je tu zakotvená, po dobu kým robot vloží na ňu podstavu. Po dokončení operácie je stanovisko 2 uvolnené a paleta smeruje k stanovisku pri poste 3. Paleta je tu zakotvená, kým sa ukončí operácia vkladania ložiska do podstavy. Po uvolnení smeruje paleta ku koncovému bodu prvého dopravníkového pásu. Po jeho dosiahnutí je manipulátorom presunutá na začiatočný bod druhého dopravníkového pásu. Po druhom dobravníkovom páse sa presunie k stanovisku pri poste 4. Tu je paleta zakotvená po dobu osadenia hriadeľky a klobúčika. Následne je paleta uvolnená a s hotovým výrobkom smeruje k stanovisku pri poste 5. Tu je paleta pozastavená po dobu odobratia hotového výrobku manipulátorom. Prázdna paleta po uvolnení smeruje ku koncovému bodu druhého pásu. Po jeho dosiahnutí je manipulátorom presunutá na začiatok prvého dopravníkového pásu. Po prvom páse sa paleta dostáva na koniec radu paliet čakajúcich na vpustenie do výroby pri stanovisku 1

Post 1

Post 1 je určený na vkladanie podstavy do výrobného procesu. Zo vstupného zásobníku je vysunutá podstava. Zásobník sa pohybuje po lineárnom vedení. Vysunutie podstavy prebieha v náhodnej polohe zásobníku. Podstava je vysunutá na šikmú plošinu. Šikmá plošina obsahuje tŕne, aby došlo aj k náhodnému pootočeniu štvorcovej podstavy. Tieto operácie slúžia na staženie podmienok. Po zošmyknutí sa podstava presúva na širokom dopravníkovom páse. Pomocou kamerového systému je snímaný dopravníkový pás. Obraz je prenášaný do aplikácie na rozpoznávanie obrazu. Táto aplikácia v reálnom čase zašle do PLC automatu aktuálnu

pozíciu a pootočenie podstavy.

Post 2

Je reprezentovaný 6-osím robotom Mitsubishi RV-2S. Po obdržaní pozície robot zdvihne podstavu a presunie ju na hlavný dopravníkový pás postu 0, kde ju osadí na paletu.

Post 3

Na poste 3 je zo vstupného zásobníku vysunuté ložisko na plošinu. Plošina sa presunie k lineárnemu potenciometru, kde prebehne meranie hrúbky ložiska. Plošina s ložiskom sa vráti na východziu pozíciu. Podľa nakonfigurovania zákazky môžu nastať 3 situácie. Nevyhovujúce ložiská sú automaticky vyhodené von zo systému. Vyhovujúce ložisko je rotačno-výsuvným manipulátorom osadené do podstavy pristavenej na poste 0. Pokiaľ nameraná hodnota hovorí o prázdnej plošine, čaká sa na doplnenie zásobníku a výroba sa na poste 3 opakuje.

Post 4

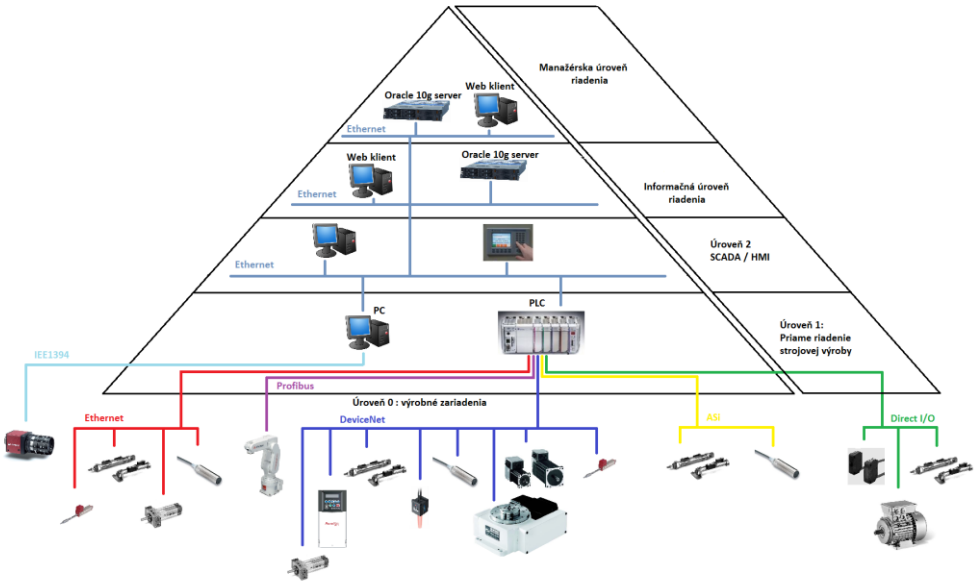
Zo vstupného zásobníka sú dávkované hriadeľky. Hriadeľky zapadajú do podstavcov na rotačnom stole. V prvej pozicii po otočení stola sa gefran snímačom zmeria výška hriadeľky. Týmto meraním sa detekuje či je hriadeľka správne vložená do podstavca. Na ďalšej pozícii rotačného stola prebieha detekcia farby hriadeľky. Pokiaľ sú namerané hodnoty z prvých dvoch pozícií nevyhovujúce hriadeľka je pneumatickým manipulátorom vyhodená zo stola. Ak hriadeľka prejde kontrolami je presunutá na poslednú - šiestu pozíciu stola. Tu ju rotačno - výsuvný manipulátor osadí do výrobku na poste 0. Pri piatej pozícii stola sa nachádza vstupný zásobník klobúčikov. Klobúčiky sú po vysunutí zo zásobníka testované na materiál pomocou indukčného snímača. Ak klobúčik nie je požadovaného materiálu, je vyhodený pomocou pneumatického piestu. Pokiaľ materiál vyhovuje objednávke, klobúčik je presunutý rotačno-výsuvným manipulátorom na rotačný stôl. Následne je rovnakým spôsobom ako hriadeľka na šiestej pozícii stola osadený na výrobok na poste 0. Po tomto kroku je výrobok hotový.

Post 5

Je tvorený výstupným skladom na hotové výrobky a trojosím manipulátorom. Sklad obsahuje 24 skladovacích pozícií. Po príchode palety s výrobkom k postu 5 je paleta ukotvená a čaká na vyskladnenie manipulátorom. Manipulátor vysunie pneumatické rameno nad paletu a štyrmi prísavkami vytvorí na povrchu výrobku podtlak. Manipulátor zdvihne hotový výrobok zasunutím pneumatického ramena. Následne sa presunie na vybranú pozíciu skladu a vysunutím ramena položí položí výrobok na bunku skladu. Vypne podtlak a vytvorí kladný tlak čím uvolní výrobok z ramena. Manipulátor sa po dokončení uloženia vracia do východzej pozície.

III. FMP V RÁMCI DISTRIBUOVANÉHO SYSTÉMU RIADENIA

Flexibilný montážny podnik ako distribuovaný systém riadenia je možné rozdeliť do jednotlivých úrovní riadenia podľa pyramídovej schémy prezentovanej spoločnosťou Rockwell Automation. Jednotlivé lokálne celky riadenia a zberu dát sú prepojené komunikačnými zbernicami. Smer distribúcie systému je od vyšších úrovní k nižším. Smer centralizácie DSR je od nižších úrovní k vyšším. Moja práca sa zaoberá riadením FMP na technologickej úrovni. Zameriavam sa preto na komunikáciu úrovní priameho riadenia strojovej výroby a úrovne výrobných zariadení. Nasledovná pyramídová schéma je priamou implementáciou schémy od Rocwell Automation na FMP.



Obr. 2 model distribuovaného systému riadenia

IV. Popis riadenia FMP

Aplikácia je tvorená v programovacom jazyku „Ladder Logic". Systém je riadený PLC automatom Allen-Bradley rady Compactlogix. Riadenie tak ako aj logické členenie linky je rozdelené pre šesť samostatných postov. Posty 1 až 5 pracujú paralelne, nezávisle jeden od druhého. Post 0 spolupracuje s každým z nich. Pre potreby globálneho riadenia systému som vytvoril podprogram „riadenie výroby". Pre externé funkcionality vizualizácie, dispečerského riadenia a zberu dát som vytvoril podprogram external a vstupno-výstupné štruktúry. Zložitejšie členy systému ako frekvenčný menič a krokové motory s lineárnym vedením majú vlastné riadiace podprogramy pre ich samotnú funkcionalitu. Po nahratí programu do PLC automatu sa prvá spustí inicializácia premenných, technologických sietí a akčných členov. Následne je riadenie odovzdané samotnej aplikácii.

V. Záver

Zkompletizoval som výrobcom navrhnutú funkcionalitu výrobného systému a vylepšil ju o detekciu a ošetrenie viacerých chybových stavov. Osvojil som si princípy riadenia frekvenčného meniča a krokových motorov s internou logikou a poznatky využil v riadení systému. Komunikáciu s vyššími vrstvami DSR som rozdelil do komunikačných štruktúr. Dosiahol som plnú plánovanú funkcionalitu systému. Riadenie som realizoval s ohľadom na variabilitu podmienok, v ktorých linka pracuje.

Poďakovanie

Táto práca bola vytvorená realizáciou projektu Rozvoj Centra informačných a komunikačných technológií pre znalostné systémy (kód ITMS projektu: 26220120030) na základe podpory operačného programu Výskum a vývoj financovaného z Európskeho fondu regionálneho rozvoja.

Literatúra

[1] Kačmár, Matej: Aplikácia systému rozpoznávania dynamických obrazov pre výrobné linky, Diplomová práca, KKUI FEI TU Košice, 2011.
[2] Papcun, Peter: Riadenie robota integrovaného v pružnej výrobnej linke, Diplomová práca, KKUI FEI TU Košice, 2011.
[3] Rocwell Automation: Powerflex 40, user manual. [online]. [2008]. Dostupné na internete: http://www.rocwellautomation.com/rocwellsoftwareoor.
[4] WEISS: Electromechanical rotary indexing tables TC-T, user manual. [online]. [2009]. Dostupné na internete: http://www.weissinstruments.com
[5] Schneider electric: ILS2D Devicenete, user manual. [online]. [2010] Dostupné na internete: http://www.schneider-electric.com
[6] J Keyence: SUPER RGB Sensor, user manual. [online]. [2011]. Dostupné na internete: http://www.keyence.com

Adaptívne prediktívne riadenie na báze analytickej a experimentálnej identifikácie riadeného systému

Štefan JAJČIŠIN, Anna JADLOVSKÁ

Katedra kybernetiky a umelej inteligencie, Fakulta elektrotechniky a informatiky,
Technická univerzita v Košiciach, Slovenská republika

stefan.jajcisin@tuke.sk, anna.jadlovska@tuke.sk

***Abstrakt* — Článok prezentuje klasický prístup pri návrhu algoritmov zovšeobecneného prediktívneho riadenia a dva odlišné spôsoby adaptácie parametrov riadiacich algoritmov v rámci riadiaceho cyklu. V článku je uvedený algoritmus adaptácie parametrov riadiaceho algoritmu v závislosti na vopred známych časových priebehoch riadiacej veličiny a taktiež algoritmus založený na experimentálne nameraných signáloch vstupných a výstupných veličín riadeného systému. V rámci článku sú porovnané výsledky simulačného riadenia nelineárneho modelu hydraulického systému pomocou klasického zovšeobecneného prediktívneho riadenia s výsledkami získanými pri použití adaptívnych algoritmov prediktívneho riadenia.**

***Kľúčové slová* — adaptívne prediktívne riadenie, zovšeobecnené prediktívne riadenie, analytická a experimentálna identifikácia, hydraulický systém**

I. ÚVOD

V článku prezentujeme základný princíp a teoretický návrh algoritmu prediktívneho riadenia pri použitom lineárnom modeli riadeného systému vo forme vstupno-výstupného opisu. Konkrétne sa zaoberáme algoritmom zovšeobecneného prediktívneho riadenia na báze CARIMA modelu [1]. Ďalej sa zaoberáme dvoma rôznymi prístupmi pre algoritmické riešenie adaptácie parametrov riadiacich algoritmov v rámci riadiacej štruktúry prediktívneho riadenia, bližšie rozoberanými v [2]. Prezentujeme prístup vychádzajúci zo známeho matematicko-fyzikálneho opisu a taktiež prístup založený na experimentálnej identifikácii parametrov modelu riadeného systému s využitím rekurzívnej metódy najmenších štvorcov. V závere článku uvádzame získané výsledky pri riadení nelineárneho simulačného modelu hydraulického systému, ktorý sa nachádza v Laboratóriu mechatronických systémov na Katedre kybernetiky a umelej inteligencie pomocou všetkých navrhnutých algoritmov prediktívneho riadenia.

II. ZOVŠEOBECNENÉ PREDIKTÍVNE RIADENIE – KLASICKÝ PRÍSTUP

Prediktívne riadenie je typické tým, že na výpočet akčného zásahu v danom kroku používa predikciu správania sa riadeného fyzikálneho systému. Vo väčšine prípadov sú predikované hodnoty príslušných veličín určované na základe lineárneho modelu riadeného systému. V tomto článku sa budeme zaoberať zovšeobecneným prediktívnym riadením (GPC), ktoré vychádza zo vstupno-výstupného opisu riadeného systému vo forme CARIMA modelu

$$A_z(z^{-1})y(k) = B_z(z^{-1})u(k-1) + \frac{C_z(z^{-1})}{\Delta}\xi(k)\,, \tag{1}$$

kde $B_z(z^{-1})$ a $A_z(z^{-1})$ sú polynómy čitateľa a menovateľa rádu m a n, $u(k)$ je vstup, $y(k)$ je výstup dynamického systému, $\xi(k)$ je chyba výstupu systému, resp. šum merania výstupu, $C_z(z^{-1})$ je polynóm a $\Delta = 1 - z^{-1}$ [1].

Metódy prediktívneho riadenia patria k moderným optimalizačným metódam. Vo všeobecnosti sa v rámci nich minimalizuje funkcionál

$$J = \sum_{i=N_1}^{N_p} \boldsymbol{Q}_e(i)\left[\hat{\boldsymbol{y}}(k+i) - \boldsymbol{w}(k+i)\right]^2 + \sum_{i=1}^{N_u} \boldsymbol{R}_u(i)\left[\Delta\boldsymbol{u}(k+i-1)\right]^2 + \sum_{i=N_1}^{N_p} \boldsymbol{S}_e(i)\left[\hat{\boldsymbol{y}}_s(k+i)\right]^2\,, \tag{2}$$

kde $\hat{\boldsymbol{y}}(k)$ je vektor hodnôt výstupu, $\boldsymbol{w}(k)$ je vektor požadovaných hodnôt výstupu, $\Delta\boldsymbol{u}(k)$ je vektor hodnôt zmeny akčného zásahu, $\hat{\boldsymbol{y}}_s(k)$ je sumátor odchýlok požadovaných a aktuálnych hodnôt výstupu, ktorého hodnota je v každom kroku k vypočítavaná na základe vzťahu

$$\hat{\boldsymbol{y}}_s(k+1) = \hat{\boldsymbol{y}}_s(k) + \boldsymbol{w}(k) - \boldsymbol{y}(k), \tag{3}$$

kde $\boldsymbol{y}(k)$ je výstup riadeného systému.

Hodnoty N_1 a N_p vo funkcionáli (2) udávajú horizont predikcie a hodnota N_u predstavuje riadiaci horizont, na ktorom sa vypočítava optimálna postupnosť hodnôt akčného zásahu $\boldsymbol{u}(k)$, pričom platí $N_u \le N_p$. Podľa [1] by hodnota N_1 mala byť najmenej $d + 1$, kde d je dopravné oneskorenie systému. V našom prípade budeme uvažovať $N_1 = 1$. Ak sa pri metódach prediktívneho riadenia používa redukcia stupňov voľnosti akčného zásahu, platí $N_u < N_p$ [2].

Parametre, ktoré môžeme v rámci funkcionálu (2) nastavovať sú:

- váhová matica odchýlok požadovaných a predikovaných hodnôt výstupov $\boldsymbol{Q}_e(i)$,
- váhová matica hodnôt akčného zásahu pre jednotlivé vstupy systému $\boldsymbol{R}_u(i)$,
- váhová matica sumátora $\boldsymbol{S}_e(i)$.

V článku budeme váhové matice $\boldsymbol{Q}_e(i)$, $\boldsymbol{R}_u(i)$ a $\boldsymbol{S}_e(i)$ považovať za diagonálne matice s hodnotami váhových koeficientov q, r, s a zároveň za konštantné na celej dĺžke oboch horizontov, takže v ďalšom sa obmedzíme na označenie $\boldsymbol{Q}_e$, $\boldsymbol{R}_u$ a $\boldsymbol{S}_e$. Tiež je potrebné poznamenať, že najviac rozšírené implementácie prediktívneho riadenia neuvažujú so sumátorom $\hat{\boldsymbol{y}}_s$, resp. matica $\boldsymbol{S}_e$ je nulová. Taktiež v tomto článku nebudeme uvažovať sumátor.

Pre algoritmy prediktívneho riadenia je typické využitie *stratégie pohyblivého horizontu* [2]. Jedná sa o vypočítanie optimálnej postupnosti akčného zásahu $\boldsymbol{u}_{opt} = \left[\boldsymbol{u}_{opt}(k) \;\; \cdots \;\; \boldsymbol{u}_{opt}(k+N_u)\right]$ na celej dĺžke riadiaceho horizontu N_u v každom kroku k, pričom na vstup systému $\boldsymbol{u}(k)$ je privedený iba prvý člen postupnosti $\boldsymbol{u}_{opt}(k)$.

Využitie stratégie pohyblivého horizontu vyžaduje poznať požadované hodnoty výstupov (referenčnú trajektóriu) $\boldsymbol{w}(k)$ na dĺžke predikčného horizontu N_p. V prípade, že referenčná trajektória nie je známa, používa sa tzv. *náhodná prechádzka*, kde $w(k+1) = w(k) + \xi(k)$, pričom $\xi(k)$ je biely šum [2].

Samotný návrh algoritmov prediktívneho riadenia je možné rozdeliť do troch fáz:

1. odvodenie prediktora na základe lineárneho modelu riadeného systému (získaného linearizáciou analytického modelu alebo ako výsledok experimentálnej identifikácie),
2. vyjadrenie gradientu $\boldsymbol{g}^T$ a Hesiánu $\boldsymbol{H}$ dosadením prediktora do funkcionálu J,
3. výpočet optimálnej postupnosti hodnôt akčného zásahu minimalizáciou funkcionálu J na základe predikcie správania sa dynamického systému bez pôsobenia budúcich hodnôt akčného zásahu, tzv. *voľnej odozvy systému*.

Vstupom do prvej fázy návrhu algoritmov prediktívneho riadenia je lineárny model riadeného fyzikálneho systému, v našom prípade vyjadrený vo forme CARIMA modelu (1). Jeho postupnou úpravou podľa [1] je možné pomocou riešenia diofantických rovníc

$$C_z = E_j \Delta A_z + z^{-j} F_j \quad \text{a} \quad B_z E_j = G_j C_z + z^{-j} \Gamma_j \tag{4}$$

vyjadriť maticový zápis prediktora v tvare

$$\hat{\boldsymbol{y}} = \boldsymbol{y}_f + \boldsymbol{G}\Delta\boldsymbol{u}, \tag{5}$$

kde $\hat{\boldsymbol{y}}$ je vektor predikovaných hodnôt výstupov, $\boldsymbol{y}_f$ je vektor hodnôt *voľnej odozvy* systému a $\boldsymbol{G}\Delta\boldsymbol{u}$ predstavuje *vynútenú odozvu* systému, pričom vychádzajúc z [1], resp. [4] je matica $\boldsymbol{G}$ tvorená koeficientami prechodovej charakteristiky systému lineárneho modelu.

Druhá fáza návrhu algoritmov prediktívneho riadenia pozostáva z dosadenia prediktora (5) do maticového tvaru funkcionálu (2) (uvažujeme váženie zmeny akčného zásahu $\Delta\boldsymbol{u}$)

$$J = \left(\hat{\boldsymbol{y}} - \boldsymbol{w}\right)^T \boldsymbol{Q}\left(\hat{\boldsymbol{y}} - \boldsymbol{w}\right) + \Delta\boldsymbol{u}^T \boldsymbol{R}\Delta\boldsymbol{u}, \tag{6}$$

kde jednotlivé vektory $\hat{\boldsymbol{y}}$, $\boldsymbol{w}$, resp. $\Delta\boldsymbol{u}$ a váhové matice $\boldsymbol{Q}$, resp. $\boldsymbol{R}$ majú rozmer v súlade s dĺžkou horizontu N_p, resp. N_u a počtom výstupov, resp. vstupov riadeného systému.

Po dosadení prediktora (5) do funckionálu (6) a následnom roznásobení získame rovnicu

$$J = \Delta\boldsymbol{u}^T\left(\boldsymbol{G}^T\boldsymbol{Q}\boldsymbol{G} + \boldsymbol{R}\right)\Delta\boldsymbol{u} + \left[\left(\boldsymbol{y}_f - \boldsymbol{w}\right)^T \boldsymbol{Q}\boldsymbol{G}\right]\Delta\boldsymbol{u} + \Delta\boldsymbol{u}^T\left[\boldsymbol{G}^T\boldsymbol{Q}\left(\boldsymbol{y}_f - \boldsymbol{w}\right)\right] + c, \tag{7}$$

ktorú je možné ďalej zjednodušiť na kvadratickú formu

$$J = \Delta\boldsymbol{u}^T\boldsymbol{H}\Delta\boldsymbol{u} + 2\boldsymbol{g}^T\Delta\boldsymbol{u} + c, \tag{8}$$

kde matica $\boldsymbol{H}$ (Hesián) a vektor $\boldsymbol{g}^T$ (gradient) majú tvar

$$\boldsymbol{H} = \boldsymbol{G}^T\boldsymbol{Q}\boldsymbol{G} + \boldsymbol{R}, \quad \boldsymbol{g}^T = \left(\boldsymbol{y}_f - \boldsymbol{w}\right)^T \boldsymbol{Q}\boldsymbol{G}. \tag{9}$$

Výhodou prediktívneho riadenia oproti klasickým postupom je možnosť zohľadniť rôzne obmedzenia (napr. akčného zásahu, zmeny akčného zásahu, výstupu) vo výpočte optimálnej postupnosti akčného zásahu numerickými metódami *kvadratického programovania*. My sme na tento účel využili funkciu *quadprog*, ktorá je súčasťou toolboxu *Optimization Toolbox* v prostredí Matlab a numericky vypočítava vektor optimálnych hodnôt $\Delta\boldsymbol{u}$ podľa vzťahu

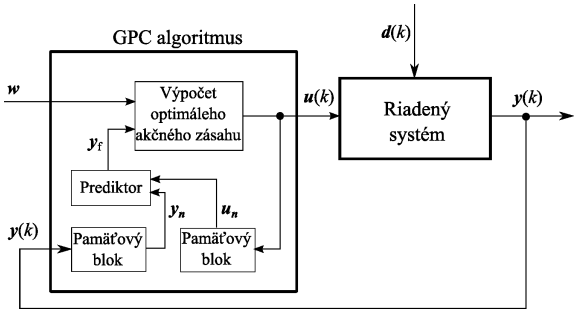
$$\min_{\boldsymbol{u}} \frac{1}{2}\Delta\boldsymbol{u}^T \boldsymbol{H}\Delta\boldsymbol{u} + \boldsymbol{g}^T\Delta\boldsymbol{u} \text{ , pri zohľadnení obmedzení } \boldsymbol{A}_{obm}\Delta\boldsymbol{u} \le \boldsymbol{b}_{obm} \text{ ,} \tag{10}$$

pričom maticu $\boldsymbol{A}_{obm}$ a vektor $\boldsymbol{b}_{obm}$ je potrebné zostaviť podľa požadovaných obmedzení. Konkrétny tvar matice $\boldsymbol{A}_{obm}$ a vektora $\boldsymbol{b}_{obm}$ pre jednotlivé obmedzenia je uvedený v [2].

V prípade, že nie je stanovená požiadavka na rešpektovanie obmedzení riadeného systému, je podľa [1], resp. [2] možné minimalizovať funkcionál analytickým spôsobom na základe vzťahu

$$\Delta\boldsymbol{u} = -\boldsymbol{H}^{-1}\boldsymbol{g} \text{ .} \tag{11}$$

Pri riadení dynamických systémov je možné algoritmus prediktívneho riadenia, ktorého návrh sme prezentovali vyššie, použiť v riadiacej štruktúre podľa Obr. 1, kde $\boldsymbol{w}$ je vektor požadovaných hodnôt výstupov na dĺžke predikčného horizontu, $\boldsymbol{u}(k)$ je akčný zásah vypočítaný GPC algoritmom, $\boldsymbol{d}(k)$ je vektor poruchových veličín a $\boldsymbol{y}(k)$ je výstup riadeného systému. Predikcia voľnej odozvy systému $\boldsymbol{y}_f$ je vypočítavaná na základe aktuálnej hodnoty výstupu $\boldsymbol{y}(k)$ a hodnôt vstupov, resp. výstupov v predchádzajúcich krokoch $\boldsymbol{u}_n$, resp. $\boldsymbol{y}_n$.



Obr. 1 Riadiaca štruktúra s GPC algoritmom

III. ADAPTÍVNE PREDIKTÍVNE RIADENIE

Použiteľnosť prezentovaného prediktívneho riadenia závisí na možnostiach vytvorenia relevantnej lineárnej aproximácie riadeného systému, ktorá by v dostatočnej miere vyjadrovala dynamiku riadeného systému v celej škále rôznych pracovných bodov. V tejto časti sa budeme zaoberať adaptívnym prediktívnym riadením, pričom pod pojmom adaptácia budeme chápať zmenu parametrov riadiaceho algoritmu (zmenu hodnôt dátových členov údajovej štruktúry reprezentujúcej riadiaci algoritmus z programátorského hľadiska) na základe aktuálneho stavu, resp. pracovného bodu riadeného systému. Budeme pritom vychádzať z analytickej a experimentálnej identifikácie riadeného systému.

A. Adaptívne prediktívne riadenie na báze analytickej identifikácie riadeného systému

Využitie analytickej identifikácie pri adaptácii parametrov riadiaceho algoritmu môže byť realizované vo forme okamžitej linearizácie riadeného systému a následnom prepočítaní príslušných parametrov, ktoré sú potrebné pre výpočet optimálneho akčného zásahu v každom kroku riadiaceho cyklu. Avšak takýto prístup je náročný na výpočtový čas a výrazne tak obmedzuje možnosti voľby periódy vzorkovania.

Iný prístup je možný vtedy, keď je známy konečný počet pracovných bodov, v ktorých by mal riadený systém pracovať, a teda je možné vopred vypočítať parametre riadiaceho algoritmu pre konkrétny pracovný bod. Na určenie množiny pracovných bodov sme navrhli a v prostredí Matlab naprogramovali funkciu so syntaxou $\boldsymbol{L}$ = *Levels(**W**,c)*, ktorá na základe citlivosti *c* zistí, aké úrovne $\boldsymbol{L}$ nadobúdajú časové priebehy usporiadané v matici $\boldsymbol{W}$. Po vypočítaní parametrov riadiaceho algoritmu pre príslušné pracovné body môžeme v rámci riadiaceho cyklu medzi nimi prepínať na základe aktuálne požadovaného pracovného bodu. Na tento účel sme naprogramovali funkciu *i=GetIndex*($\boldsymbol{L}$,$\boldsymbol{v}$), v rámci ktorej je realizovaný algoritmus hľadania najkratšej vzdialenosti vektora aktuálneho pracovného bodu $\boldsymbol{v}$ od jednotlivých pracovných bodov $\boldsymbol{L}$. Funkcia vracia poradové číslo najbližšieho *i*-tého pracovného bodu podľa metriky

$$d(\boldsymbol{L}_i, \boldsymbol{v}) = \sqrt{\sum_{j=1}^{nO} (\boldsymbol{L}_{i,j} - \boldsymbol{v}_j)^2}, \tag{12}$$

kde *nO* je dĺžka vektora $\boldsymbol{v}$, resp. $\boldsymbol{L}_i$, a teda aj počet výstupov riadeného systému.

B. Adaptívne prediktívne riadenie na báze experimentálnej identifikácie riadeného systému

Podstata tohto prístupu k adaptácii algoritmov prediktívneho riadenia spočíva v určení lineárneho modelu riadeného systému a jeho následnom použití pri výpočte parametrov riadiaceho algoritmu v rámci periódy vzorkovania. Výpočet parametrov lineárneho modelu budeme realizovať experimentálnou identifikáciou s využitím rekurzívnej metódy najmenších štvorcov (RMNŠ) podľa [5]. Vstupnými údajmi pre RMNŠ v danom kroku sú hodnoty časových priebehov vstupov a výstupov riadeného systému v *m*, resp. *n* predchádzajúcich vzorkách. Výsledkom sú koeficienty čitateľa b_i a menovateľa a_i prenosovej funkcie

$$F(z^{-1}) = \frac{b_1 z^{-1} + b_2 z^{-2} + ... + b_m z^{-m}}{1 + a_1 z^{-1} + a_2 z^{-2} + ... + a_n z^{-n}}, \tag{13}$$

ktorá reprezentuje lineárny model riadeného systému potrebný pri výpočte parametrov algoritmov prediktívneho riadenia.

Podstata RMNŠ spočíva v korekcii predchádzajúcich hodnôt koeficientov b_i a a_i na základe aktuálneho výstupu systému, ktorý je odozvou na predchádzajúce hodnoty vstupu a výstupu. Ak $\boldsymbol{\theta} = \begin{bmatrix} a_1 & a_2 & \cdots & a_n & b_1 & b_2 & \cdots & b_m \end{bmatrix}^T$ je vektor neznámych koeficientov prenosovej funkcie (13) a $\boldsymbol{\varphi}(k) = \begin{bmatrix} y(k-1) & y(k-2) & \cdots & y(k-n) & u(k-1) & u(k-2) & \cdots & u(k-m) \end{bmatrix}^T$ je vektor predchádzajúcich hodnôt vstupov a výstupov systému, môžeme výpočet nových hodnôt koeficientov b_i a a_i vyjadriť prostredníctvom vzťahu

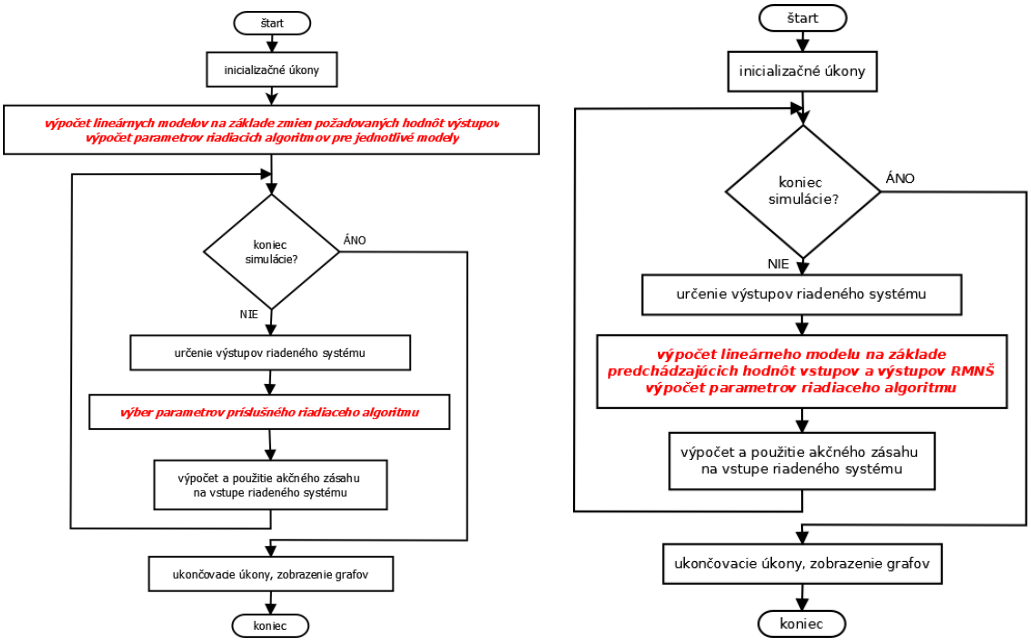
$$\underbrace{\boldsymbol{\theta}(k+1)}_{\substack{\textit{nové hodnoty} \\ \text{parametrov}}} = \underbrace{\boldsymbol{\theta}(k)}_{\substack{\textit{predchádzajúce} \\ \textit{hodnoty parametrov}}} + \underbrace{\boldsymbol{\tau}(k+1)}_{\textit{korekčný vektor}} \underbrace{\left[y(k+1) - \boldsymbol{\varphi}^T(k+1)\boldsymbol{\theta}(k) \right]}_{\textit{predikčná chyba}}, \tag{14}$$

kde $\boldsymbol{\tau}(k+1) = \boldsymbol{P}(k)\boldsymbol{\varphi}(k+1)\left[\sigma(k+1) + \boldsymbol{\varphi}^T(k+1)\boldsymbol{P}(k)\boldsymbol{\varphi}(k+1) \right]^{-1}$, pričom σ je faktor zabúdania a $\boldsymbol{P}$ je kovariančná matica, ktorej hodnoty sú v každom kroku dané vzorcom

$$\boldsymbol{P}(k+1) = \frac{1}{\sigma(k+1)} \left\{ \boldsymbol{P}(k) - \boldsymbol{P}(k)\boldsymbol{\varphi}(k+1)\left[\sigma(k+1) + \boldsymbol{\varphi}^T(k+1)\boldsymbol{P}(k)\boldsymbol{\varphi}(k+1) \right]^{-1} \boldsymbol{\varphi}^T(k+1)\boldsymbol{P}(k) \right\}. \tag{15}$$

Algoritmus RMNŠ sme naprogramovali v podobe m-funkcie [$\boldsymbol{B}_v$, $\boldsymbol{A}_v$] = *RLS*($\boldsymbol{y}_i$, $\boldsymbol{u}_i$, $\boldsymbol{N}_{ab}$, *p*, *FZ*), kde $\boldsymbol{y}_i$, $\boldsymbol{u}_i$ sú vektory predchádzajúcich hodnôt vstupov a výstupov, $\boldsymbol{N}_{ab}$ je dvojprvkový vektor s požadovanými hodnotami rádu menovateľa a čitateľa prenosu (13), parameter *p* slúži pre inicializáciu kovariančnej matice v podobe jednotkovej matice na hlavnej diagonále a *FZ* je hodnota faktora zabúdania. Funkcia *RLS* vracia vektory koeficienty čitateľa $\boldsymbol{B}_v$ a menovateľa $\boldsymbol{A}_v$.

Riadiaca štruktúra na Obr. 1 je platná aj v prípade algoritmov adaptívneho prediktívneho riadenia. Avšak jej algoritmizáciu pri riadení dynamických systémov je potrebné realizovať v súlade vývojovými diagramami uvedenými na Obr.2.

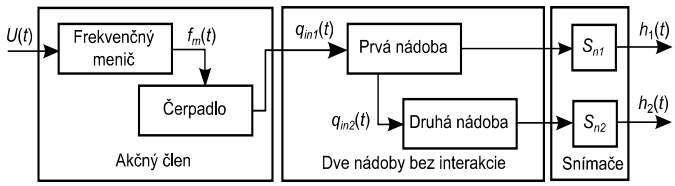


Obr. 2 Vývojové diagramy algoritmizácie riadiacej štruktúry s adaptívnym GPC algoritmom na báze analytickej (vľavo) a experimentálnej identifikácie (vpravo)

Najvýraznejší rozdiel oboch prezentovaných prístupov k adaptácii algoritmov prediktívneho riadenia spočíva vo výpočtovej náročnosti algoritmov v rámci jednej periódy vzorkovania. Použitie prístupu s experimentálnou identifikáciou kladie vyššie nároky na voľbu periódy vzorkovania, keďže v rámci jedného riadiaceho cyklu je potrebné odhadnúť parametre lineárneho modelu riadeného systému a zároveň kompletne prepočítať parametre riadiaceho algoritmu. Na druhej strane, veľkou výhodou tohto prístupu je jeho použiteľnosť v prípadoch, kedy nie je možné získať lineárny model riadeného systému matematicko-fyzikálnym opisom. Tento prístup je efektívnejší aj pri riadení dynamických systémov s premenlivými parametrami, ktorých zmenu nie je možné zahrnúť vopred do matematicko-fyzikálneho opisu, teda v rámci analytickej identifikácie.

IV. VÝSLEDKY RIADENIA HYDRAULICKÉHO SYSTÉMU S ADAPTÍVNYM GPC ALGORITMOM

V tejto časti uvádzame výsledky simulačného riadenia nelineárneho modelu hydraulického systému pomocou klasického a adaptívnych algoritmov riadenia prezentovaných v predchádzajúcej časti. Detailný popis modelu hydraulického systému, jeho hardvérová konfigurácia a popis fyzikálnych veličín a parametrov je uvedený v [6] a [7]. V tomto článku ho predstavíme len veľmi stručne. Jedná sa o laboratórny model hydraulického systému dvoch nádob, ktorý sa nachádza v Laboratóriu mechatronických systémov na Katedre kybernetiky a umelej inteligencie. Kvapalina je privádzaná do prvej nádoby zo zásobníka prostredníctvom čerpadla, ktoré je priamo ovládané frekvenciou frekvenčného meniča $f_m(t)$, resp. riadiacim napätím $U(t)$ v rozsahu 0 – 10V. Kvapalina vytekajúca z prvej nádoby preteká hadičkou do druhej nádoby, z ktorej sa opätovne vracia do zásobníka. Výšky hladín kvapaliny v oboch nádobách $h_1(t)$ a $h_2(t)$ sú merané kapacitnými snímačmi S_{n1} a S_{n2} a ich hodnoty sa pohybujú v rozsahu 0 – 0,3m. Zo systémového hľadiska môžeme tento systém znázorniť pomocou Obr. 3, kde $q_{in1}(t)$ a $q_{in2}(t)$ sú vstupné prítoky do nádob.



Obr. 3 Systémový pohľad na hydraulický systém

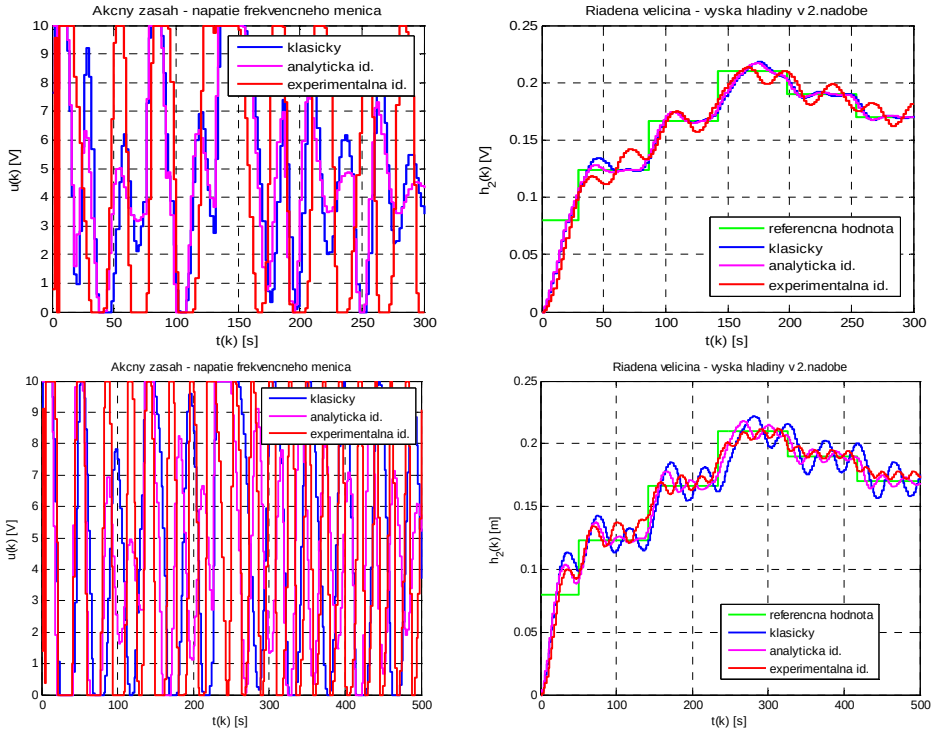
Matematicko-fyzikálny opis hydraulického systému tvoria nelineárne diferenciálne rovnice

$$\frac{dh_1(t)}{dt} = \frac{1}{S}\left(k_p U(t) - S_{v1}\sqrt{2gh_1(t)}\right)$$
$$\frac{dh_2(t)}{dt} = \frac{1}{S}\left(S_{v1}\sqrt{2gh_1(t)} - S_{v2}\sqrt{2gh_2(t)}\right) \quad , \qquad (16)$$

kde S je prierez nádob, k_p je konštanta čerpadla vyjadrujúca vzťah medzi vstupným napätím $U(t)$ a prítokom $q_{in1}(t)$, g je gravitačné zrýchlenie a S_{v1}, S_{v2} sú prierezy výstupných ventilov nádob.

Na základe nelineárnych diferenciálnych rovníc (16) sme zostavili simulačný model v podobe funkcie jazyka Matlab, ktorý predstavuje riadený systém. Za referenčný signál sme zvolili skokovú funkciu s niekoľkými úrovňami. Linearizáciou rovníc (16) pre jednotlivé požadované úrovne (pracovné body), sme vytvorili množinu lineárnych modelov. Výpočet parametrov riadiacich algoritmov, či už s využitím vopred určených lineárnych modelov alebo na báze experimentálnej identifikácie sme realizovali s hodnotami váhových koeficientov a jednotlivých horizontov funkcionálu (2), resp. (6): váhové koeficienty q = 1000; r = 0,01; s = 0; horizont predikcie N_p = 8 a 5; riadiaci horizont N_u = 2. Pri výpočte akčného zásahu algoritmom prediktívneho riadenia sme uvažovali s vyššie uvedenými obmedzeniami vstupov a výstupov riadeného systému.

Na Obr. 4 sú porovnané výsledky simulačného riadenia hydraulického systému pomocou klasického GPC algoritmu s výsledkami získanými pomocou oboch prezentovaných adaptívnych algoritmov. Perióda vzorkovania bola T_{vz} = 2s, v hornej časti obrázka sú výsledky, kde N_p = 8 vzoriek, v dolnej časti pri N_p = 5 vzoriek. Z Obr. 4 je vidieť, že pri dostatočne dlhom horizonte predikcie sú výsledky pomocou klasického GPC algoritmu (navrhnutého z lineárneho modelu pre jeden reprezentatívny pracovný bod) takmer identické s výsledkami pri použitom adaptívnom algoritme na báze analytickej identifikácie (navrhnutého pre viacero príslušných pracovných bodov). Avšak použitie klasického GPC algoritmu pri kratšom horizonte predikcie spôsobuje veľké prekmity riadenej veličiny, kým adaptívne algoritmy (hoci tiež mierne kmitavo) udržujú riadenú veličinu v blízkosti referenčnej hodnoty. Z porovnania analytického a experimentálneho prístupu sa dá usudzovať, že použitie experimentálnej identifikácie je menej výhodné. Avšak, ako už bolo spomenuté, veľkou výhodou tohto prístupu je jeho použiteľnosť v prípadoch, kedy nie je možné analyticky odvodiť model riadeného systému a k dispozícii sú iba experimentálne namerané údaje reprezentujúce dynamiku fyzikálneho systému.



Obr. 4 Výsledky simulačného riadenia nelineárneho hydraulického systému pomocou rozoberaných algoritmov

V. ZÁVER

Cieľom článku bolo prezentovať prístupy algoritmizácie adaptívneho prediktívneho riadenia a porovnať získané výsledky s použitím klasického zovšeobecneného prediktívneho riadenia. Vzhľadom na to, že všetky uvedené prístupy sú založené na aproximácii dynamiky riadeného systému lineárnym modelom, ako ďalšiu možnosť pri návrhu adaptívnych algoritmov vidíme použitie metód umelej inteligencie, konkrétne použitie dopredných neurónových sietí v prediktívnom riadení nelineárnych systémov.

POĎAKOVANIE

Tento článok bol vytvorený realizáciou projektu Rozvoj Centra informačných a komunikačných technológií pre znalostné systémy (ITMS kód: 26220120030) na základe podpory operačného programu Výskum a vývoj financovaného z Európskeho fondu regionálneho rozvoja.

LITERATÚRA

[1] FIKAR, M.: Predictive Control – An Introduction. Bratislava: Slovenská technická univerzita - FCHPT, 1999.
[2] JAJČIŠIN, Š.: Aplikácia moderných metód v prediktívnom riadení modelov fyzikálnych systémov. Písomná práca k dizertačnej skúške. Košice: TU-FEI, 2012.
[3] ROUBAL, J. - HAVLENA, V.: Range Control MPC Approach For Two-Dimensional System, Proceedings of the 16th IFAC World Congress, Volume # 16, Part# 1, 2005.
[4] CLARKE, D.W. – MOHDANI, C. – TUFFS, P.S.: Generalized Predictive Control, Part 1 and 2. In Automatica, Vol 23, No.2, 137 – 160, 1987.
[5] FIKAR, M. – MIKLEŠ, J.: Identifikácia systémov. CHTF STU Bratislava, 1998, ISBN N 80-227-1177-2.
[6] JAJČIŠIN, Š.: Verification of Control Algorithms with DDE Communication on Real Hydraulic System, in: SCYR 2011:11th Scientific Conference of Young Researchers of Faculty of Electrical Engineering and Informatics Technical University of Košice, pp. 388-391, 2011, ISBN 978-80-553-0644-5.
[7] JAJČIŠIN, Š. - JADLOVSKÁ, A.: Riadenie laboratórneho modelu hydraulického systému, in: Electroscope – online časopis pro elektrotechniku, ročník 2011, č. III. ISSN 1802-4564, Dostupné na internete: http://147.228.94.30/index.php?option=com_content&view=article&id=280:riadenie-laboratorneho-modelu-hydraulickeho-systemu-&catid=34:cislo-32011-&Itemid=48.

Efektívne využívanie zdrojov vo výrobe pomocou procesného modelovania

Slavomíra JURČÍKOVÁ, František BABIČ

Katedra kybernetiky a umelej inteligencie, Fakulta elektrotechniky a informatiky,
Technická univerzita v Košiciach, Slovenská republika

slavomira.jurcikova@student.tuke.sk, frantisek.babic@tuke.sk

***Abstrakt* — Cieľom implementácie procesného modelovania do prostredia výrobnej firmy je zefektívniť realizáciu vykonávaných procesov za účelom zvýšenia zisku, zníženia výrobných nákladov a zavádzania nových technologických riešení. V tomto prípade ide o výrobný závod Whirpool Slovakia spol. s.r.o. so sídlom v Poprade, ktorý sa primárne venuje výrobe automatických pračiek. Tento výrobný proces bol podrobený analýze, na základe ktorej boli prostredníctvom nástroja pre procesné modelovanie ADONIS vytvorené príslušné modely. Pomocou týchto modelov firma získala prehľadnú a jednoducho pochopiteľnú vizualizáciu, ktorú môže podnik ďalej využívať na zlepšovanie procesov a zavádzanie nových prvkov do výroby.**

***Kľúčové slová* — výrobný proces, ADONIS, procesný model**

I. Úvod

V dnešnej dobe je zmena dôležitou súčasťou každého podniku, ktorá sa neobjavuje raz za čas, ale neustále. Technológie sa neustále vyvíjajú a snažia sa poskytnúť lepšie možnosti pre pochopenie podnikových procesov a integráciu údajov pre ďalšie použitie ako napr. podpora strategického rozhodovania, implementácia zlepšení do výroby s cieľom optimalizácie, zvyšovania zisku a znižovania nákladov. Optimalizácia môže mať viacero foriem, v tomto prípade bude aplikované procesné modelovanie na vybraný výrobný proces v reálnej firme.

Modelovania výrobných procesov v podniku vedie k zvýšeniu efektívnosti a kvality. Napríklad, ak sa nájde v podniku výrobná linka, ktorá nefunguje správne, tak je podnik stratový, má vyššie náklady a nižší zisk. Aby bolo možné navrhnúť čo najlepšie riešenia, je dôležité poznať všetky procesy, ktoré sa podieľajú na výrobe.

V oblasti procesného modelovania existuje viacero štandardov, ako vytvoriť a reprezentovať procesné modely. Najznámejším a najpoužívanejším je štandard BPMN (Business Process Modeling Notation[1]), ktorého cieľom je predovšetkým zrozumiteľnosť popisu procesov pre človeka. Pri vytváraní modelov bola použitá jeho aktuálna verzia 2.0, ktorá oproti predchádzajúcim verziám podporuje sémantiku, spresňuje udalosti zloženia a vzťahov, rozširuje definíciu ľudských interakcií a definuje choreografiu modelu.

Tento článok pozostáva z troch hlavných častí, kde prvá popisuje motiváciu; druhá je venovaná konkrétnemu výrobného procesu, ktorý bude modelovaný prostredníctvom nástroja ADONIS a v závere je zhodnotený celý postup a dosiahnuté výsledky.

II. Výrobný Proces

A. Popis firmy

Spoločnosť Whirlpool Slovakia spol. s.r.o. patrí do korporácie firiem Whirlpool Corporation. Tento závod exportuje svoje výrobky do viac ako 170 krajín sveta a predstavuje najväčší závod na výrobu automatických práčok plnených zhora a spredu. Z celkového počtu vyrobených práčok zostávajú na Slovensku len 3%, zvyšný objem produkcie putuje do Európy (Francúzsko, Nemecko a Taliansko) a krajín ako Rusko, Argentína alebo Hongkong.

Tento podnik podporuje snahy o neustále zdokonaľovanie výroby pre zlepšenie kvality

[1] http://www.bpmn.org/

výrobkov. Z tohto dôvodu bol priamo v podniku vybudovaný „Úsek vývoja a inovácií“, ktorý sa stal veľmi dôležitou časťou podniku.

B. Využívanie zdrojov v procese výroby

Na základe konzultácii so zodpovednými pracovníkmi na jednotlivých úsekoch výrobného procesu automatickej práčky bol identifikovaný hlavný problém, ktorý znižuje celkovú efektivitu výroby a to objem stratového materiálu spôsobený nevyhovujúcim stavom niektorých strojov, aj keď sa podnik snaží o ich zlepšenie, opravu, vylepšenie alebo nahrádza ich novšími strojmi. Tento fakt potom spôsobuje ďalšie problémy ako nespoľahlivú výrobu, vysoké náklady a vysoký počet nepodarkov. Dokumentuje to aj hodnota ukazovateľa OEE, ktorý meria celkovú efektivitu výrobných zariadení, výrobných liniek, technologických celkov, alebo aj celých výrobných závodov. Cieľom firmy je v konečnom dôsledku dosiahnuť jeho hodnotu aspoň 75%, k čomu prispeje aj aplikácia procesného modelovania na hlavný výrobný proces s cieľom optimalizácie. Ďalšou požiadavkou bolo zlepšenie procesov a zníženie nákladov vynaložených na výrobu jedného hotového výrobku skladajúceho sa z niekoľkých dielov. Tieto diely sú vyrábané v osobitných podprocesoch na rôznych linkách a každý podproces výroby príslušného dielu je niečím charakteristický. Aj keď firma používa v rámci týchto procesov kontrolné mechanizmy, ktoré kontrolujú diely a celkový výrobok, stále sú existujúce náklady vysoké. Preto firma sa snaží aspoň o mierne zníženie nákladov.

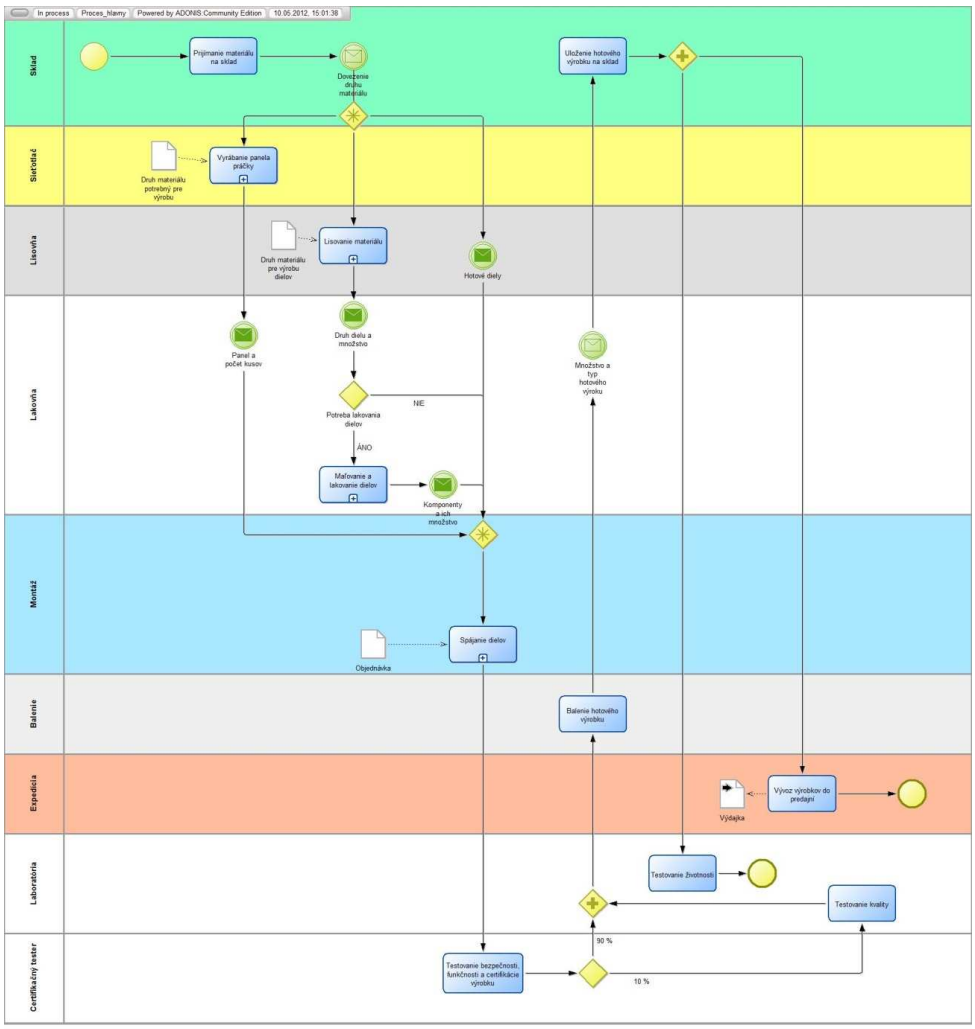
C. Procesné modely

Pre návrh a vytvorenie príslušných procesných modelov bol vybraný modelovací nástroj ADONIS, ktorý je vhodný pre skúsenejších používateľov, je na pohľad zložitejší ako napríklad BizAgi, ale jednoduchšie sa spracovávajú príslušné procesy. ADONIS ponúka aj možnosť vytvorenia tzv. aplikačných scenárov ako organizačný vývoj, riadenie kvality, riadenie rizika, procesné smernice, atď.

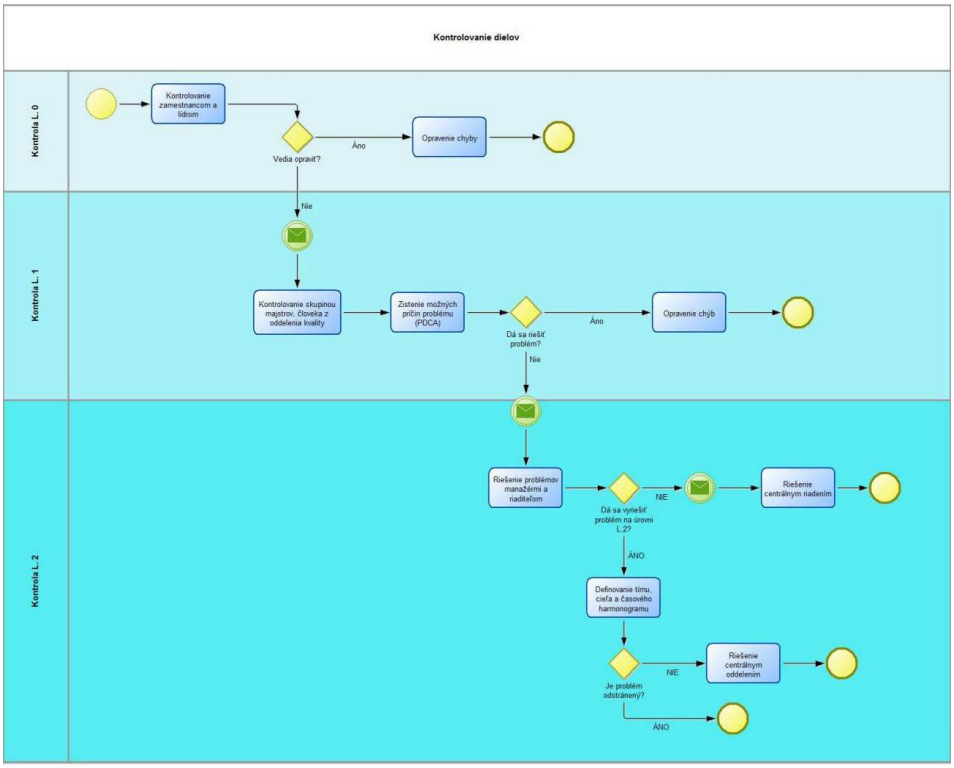
Hlavný model výroby automatickej práčky (Obr. 1) sa začína objednávkou od zákazníka, ktorá príde na oddelenie OTD, kde sa spracuje a zadá nielen do plánovacieho systému ale aj do systému pre správu dodávateľského reťazca v podniku. Dôležitým krokom je priradiť prioritu každej objednávke, kde najvyššiu majú práve tie práčky, ktoré idú priamo zákazníkovi.

Keď plánovač výroby plán týždennej výroby podľa typu práčky, expeduje sa zo skladu druh a potrebné množstvo materiálu na vyhotovenie objednávky. Súčasne pokračuje materiál na linky: Sieťotlač, Lisovňa, Montáž. Úlohou sieťotlače je vyrobiť ovládací panel automatickej práčky. Pomocou emailu plánovač rozpošle informácie o vyrábanom modeli a zodpovední pracovníci (zásobovači) si v systéme nájdu potrebné informácie o vyrábanom modeli. Na základe rozboru dodajú potrebný materiál na sieťotlač. Následne sa začína podproces výroby panela automatickej práčky. V lisovni sa vyrábajú konkrétne diely, potrebné na kompletizáciu práčky. Opäť plánovač rozpošle všetky potrebné informácie, a na ich základe zásobovači dodajú materiál, osobitne pre každú výrobnú linku. Nie všetky diely sa montujú priamo v závode v Poprade. Tieto sa dovážajú od dodávateľov a následne vstupujú do podprocesu Predmontáž spájania dielov, kde sa spájajú jednotlivé diely do celkov zjednodušujúcich montáž. Na sieťotlači sa kontroluje počet vyrobených kusov panelov a nepodarkov (Obr. 2), ktoré následne idú do odpadu. Z lisovania materiálu každá linka odosiela informácie o druhu vyrobeného dielu a o jeho počte. Keď sú vyrobené potrebné diely, časť dielov pokračuje na podproces lakovania dielov. Diely, ktoré nepotrebujú lakovanie, pokračujú ďalej v procese. Pri spájaní dielov sa postupuje presne podľa objednávky, teda všetky pripravené diely sa spájajú podľa technologického postupu. Rôzne typy práčok majú rôzne diely, preto je potrebné dávať pozor, aký diel sa montuje. Tiež je nutné dbať na správne umiestnenie dielov, aby sa zamedzilo vzniku nepodarkov.

Vyrobená práčka postupuje na certifikačný test, prostredníctvom ktorého sa overuje bezpečnosť a certifikácia hotového výrobku. Po úspešnom absolvovaní cca 10 % z celkovej produkcie pokračuje na testovanie kvality, zvyšných 90 % pokračuje na balenie. Balenie výrobku pozostáva z položenia ochranných prvkov na prepravu, ktoré zabraňujú poškodeniu automatickej práčky pri preprave. Po dodaní na expedíciu sa odpáskuje určitý počet automatických práčok, ktoré idú na test životnosti. Zo skladu hotových výrobkov sa práčky, ktoré neodchádzajú na test životnosti, nakladajú do vagónov, vlakov a kamiónov, ktoré odvážajú výrobky do predajní na celom svete. Zo skladu vychádza výdajka o práčkach, ich množstve a type. Expedíciou sa celý výrobný proces automatickej práčky končí. Všetky informácie o výrobe a expedícií sa evidujú.



Obr. 1 Hlavný výrobný proces automatickej práčky.



Obr. 2 Podproces kontroly jednotlivých dielov.

III. ZÁVER

Výrobný závod Whirlpool Slovakia spol. s.r.o. so sídlom v Poprade predstavuje inovatívny závod, ktorý neustále vylepšuje svoje procesy a snaží sa mať prehľad o nových technológiách, ktoré sa týkajú nielen samotnej výroby ale aj celkového pohľadu na manažment podniku. Závod je takisto ochotný spolupracovať vo viacerých oblastiach a projektoch, čo mu umožňuje získavať ďalšie znalosti a poznatky pre zlepšenie pozície na trhu.

Prostredníctvom implementácie a adaptácie procesného modelovania vstúpili do podniku ďalšie nové technológie a systémy. V spolupráci s Úsekom vývoja a inovácií boli zdokumentované jednotlivé úseky výroby, identifikované závislosti a problematické miesta, ktoré spôsobujú buď zdržanie vo výrobe alebo produkciu nepodarkov znižujúcu celkovú efektívnosť výrobného procesu. Prostredníctvom vytvorených modelov bol zlepšený celkový systém kontroly vo výrobe, ktorého cieľom je práve minimalizovať objem nekvalitných alebo nepodarených výrobkov. Prezentované modely v tomto článku predstavujú len časť celkovej dokumentácie, ktorá bude slúžiť firme ako manuál pri rozhodovaní a pri zlepšovaní dôležitých procesov. Podľa nich už výrobný závod vie, čo sa vyrába na jednotlivých linkách, aká linka je dôležitá a pod. Tieto modely by mali taktiež pomôcť pri lepšej informovanosti pracovníkov na všetkých úrovniach podniku.

Netreba zabúdať, že procesné riadenie závodu nie je jednorazová záležitosť a je potrebné neustále procesy sledovať a pracovať na ich zlepšení. V tomto prípade je možné použiť aj vhodné simulačné metódy, ktoré pomôžu odhaliť ďalšie skryté pozitívne alebo negatívne aspekty sledovaných procesov.

POĎAKOVANIE

Tento príspevok vznikol za podpory KEGA projektu č. 065TUKE-4/2011 (50%) a realizáciou projektu Rozvoj Centra informačných a komunikačných technológií pre znalostné systémy (kód ITMS projektu: 26220120030) na základe podpory operačného programu Výskum a vývoj financovaného z Európskeho fondu regionálneho rozvoja (50%).

Automatická tvorba korpusov pre potreby analýzy sentimentu

Peter KONCZ, Ján PARALIČ

Katedra kybernetiky a umelej inteligencie, Fakulta elektrotechniky a informatiky,
Technická univerzita v Košiciach, Slovenská republika

peter.koncz@tuke.sk, jan.paralic@tuke.sk

***Abstrakt* — Analýze sentimentu, umožňujúcej automatickú kvantifikáciu slovných hodnotení, je v súčasnosti venovaná veľká pozornosť. Väčšina existujúcich metód analýzy sentimentu je však závislá na interakcii s človekom. Táto znalosť je reprezentovaná zväčša v podobe manuálne vytváraných slovníkov slov alebo manuálne anotovaných textov. V práci je popisovaná metóda využívajúca automatickú tvorbu korpusu hodnotení, ktorý je následne použitý pri tvorbe klasifikátorov založených na metóde podporných vektorov. Korpus je automaticky prispôsobovaný obsahu dokumentov, na ktoré má byť klasifikátor aplikovaný. Výsledky experimentov ukazujú, že navrhované riešenie poskytuje relatívne presné odhady emočnej orientácie textu, hoci zvýšenie presnosti klasifikácie zohľadnením tematickej podobnosti dokumentov sa nepodarilo potvrdiť.**

***Kľúčové slová* — analýza sentimentu, dolovanie v textoch, identifikácia témy, strojové učenie.**

I. ÚVOD

Analýza sentimentu je oblasťou venovanou automatickej kvantifikácii kvalitatívnych hodnotení [1]. Jej popularita vyplýva predovšetkým z veľkého množstva on-line hodnotení, obsahujúcich cenné informácie z pohľadu rozhodovania. V súčasnosti existuje veľké množstvo prác, ktoré sa venujú metódam analýzy sentimentu. Prehľad týchto prác je možné nájsť v práci Koncz a Paralič[2]. Väčšina metód popisovaných v týchto prácach je však založená na relatívne výraznej asistencii človeka. Táto práca je preto venovaná metóde automatickej tvorby korpusov pre potreby analýzy sentimentu. Riešenie je postavené na extrakcii on-line hodnotení a ich následnom použití pri automatickej tvorbe doménovo relevantných korpusov. V nasledujúcich častiach práce sú najskôr popísané súvisiace práce. Následne je uvedené navrhované riešenie, za ktorým nasleduje jeho experimentálne overenie. Na základe získaných výsledkov sú v závere sumarizované zistenia a formulované ďalšie výskumné ciele.

II. SÚVISIACE PRÁCE

Existujúce metódy analýzy sentimentu zvyknú byť delené na dve skupiny, pričom obe skupiny metód sú závislé na zásahu človeka. Prvá skupina metód, predstavovaná tzv. na slovníkoch založenými metódami resp. exogénnymi metódami [2], je závislá na externých znalostiach vyjadrených v podobe slovníkov a pravidiel pre prácu s nimi. Slovníky pozitívnych, negatívnych, prípadne intenzitu či orientáciu meniacich slov sú často vytvárané manuálne. V snahe obmedziť závislosť na manuálnej tvorbe slovníkov existuje niekoľko prác venovaných možnostiam automatickej tvorby slovníkov. Hu a Liu[3] v rámci identifikácie orientácie hodnotení použili slovníky pozitívnych, negatívnych a negujúcich slov. Pozitívne a negatívne slová boli získané z WordNetu použitím malej skupiny manuálne anotovaných prídavných mien. Pri generovaní slov sa využil predpoklad, že slová v rámci jedného synsetu (synonymický rad) majú rovnakú orientáciu. Uvedená metóda bola použitá aj v práci Miao a kol. [4].Kamps a kol. [5] popisujú metódu pre určenie orientácie prídavných mien na základe ich vzdialenosti od slov dobrý (good) a zlý (bad) v grafe vytvorenom na základe WordNetu. V práci Kim a Hovy[6] bol WordNet použitý na rozšírenie množiny emocionálnych slov prostredníctvom ich synoným a antoným. Kvalita slovníkov pre analýzu sentimentu je však rozhodujúcou z pohľadu dosahovaných presností a preto sú manuálne vytvárané slovníky vhodnejšie. Na druhej strane závisí presnosť metód využívajúcich slovníky aj na veľkosti slovníkov a z tohto pohľadu poskytujú metódy automatickej tvorby slovníkov jednoduchý spôsob získania nových slov.

Druhou skupinou metód sú metódy založené na strojovom učení resp. endogénne metódy [2]. Tieto metódy využívajú manuálne anotované korpusy, na základe ktorých sú následne

trénované klasifikátory identifikujúce subjektívnu orientáciu textu. V mnohých prípadoch sú manuálne anotované korpusy vytvárané na základe on-line hodnotení, ich zaradeniu do korpusov však predchádzajú viaceré kroky predspracovania, v rámci ktorých je vhodnosť hodnotení overovaná anotátormi. V rámci tejto skupiny metód boli použité viaceré metódy strojového učenia, v rámci ktorých sa ako najvhodnejšia ukazuje metóda podporných vektorov [1]. Táto metóda sa ako najvhodnejšia ukázala aj v rámci nami realizovaných prác [2, 7]. Použitím tzv. jadrových funkcií (kernelfunkcions) je pritom túto metódu možné aplikovať aj na relatívne veľkých vzorkách údajov. V rámci existujúceho výskumu je relatívne malá pozornosť venovaná možnosti aplikovania vytvorených modelov na hodnotenia z iných domén. Možnosti prenosu klasifikátorov medzi doménami sa vo svojej práci venovali aj Tan a Zhang[8], ktorí vo svojej práci overovali metódu výberu atribútov pre úlohy analýzy sentimentu. Použili vzorku tvorenú 1021 hodnoteniami z oblasti vzdelávania, filmu a domácnosti. Klasifikátory dosahovali podstatne nižšie presnosti mimo domény pre ktorú boli trénované.

III. NAVRHOVANÉ RIEŠENIE

Vyššie popisované metódy analýzy sentimentu, ako bolo uvedené, vo veľkej miere závisia na manuálnej anotácii korpusov, ako aj na tvorbe slovníkov a pravidiel ich aplikácie. Jedným z cieľov nami navrhovaného riešenia je preto minimalizácia množstva manuálnej práce.

Ďalšou charakteristikou mnohých metód analýzy sentimentu je ich relatívne vysoká výpočtová náročnosť. Vyššia výpočtová náročnosť je umožňovaná relatívne malými veľkosťami korpusov používaných pri overovaní týchto metód, ktoré obsahujú nanajvýš niekoľko tisíc hodnotení. V rámci navrhovanej metódy je preto pozornosť sústredená aj na možnosť jej aplikácie na veľkých množstvách hodnotení.

Navrhovaná metóda automatickej konštrukcie korpusov pozostáva z nasledujúcich krokov:

1) *Získanie stránok s on-line hodnoteniami*. Štruktúra stránok obsahujúcich hodnotenia produktov a služieb je relatívne podobná. Väčšina z nich obsahuje textové hodnotenie a bodové hodnotenie, ktoré je zväčša realizované v podobe päťbodovej škály. Tieto stránky je možné stiahnuť použitím vhodného crawlera. Z pohľadu používateľa je požadované definovanie počiatočnej stránky a pravidiel prehľadávania a ukladania stránok. Z pohľadu výpočtovej náročnosti je crawlovanie možné efektívne paralelizovať.

2) *Extrahovanie hodnotení*. K extrahovaniu textových a bodových hodnotení je možné použiť XPath výrazy, pomocou ktorých sa najskôr extrahujú jednotlivé hodnotenia a z nich sa následne extrahujú textové a bodové hodnotenia. Po extrahovaní obsahu a hodnotení je potrebné spraviť predspracovania textových hodnotení a diskretizáciu bodových hodnotení. Z pohľadu používateľa stačí na tieto účely definovať tri XPath výrazy, hoci extrahované môžu byť aj ďalšie atribúty hodnotení. Podobne aj tento krok je možné efektívne paralelizovať.

3) *Výber tematicky relevantných hodnotení*. Vytvorená vzorka údajov môže byť priamo použitá pre trénovanie klasifikátorov. Problémom zostáva relatívne vysoká doménová závislosť klasifikátorov. Ďalším možným krokom pri príprave korpusu je preto identifikácia témy korpusu, ktorá by mala byť rovnaká ako téma hodnotení, na ktoré má byť klasifikátor aplikovaný. Z toho dôvodu sme v realizovaných experimentoch použili automatické zoradenie hodnotení v trénovacej vzorke údajov podľa miery ich relevantnosti voči tematickej doméne hodnotení v testovacej vzorke údajov.

IV. EXPERIMENT

A. Vzorka

Pre trénovanie bola použitá vzorka údajov pochádzajúca z portálu Reviewcenter[1]. Z pôvodnej množiny hodnotení bolo vybraných 17422 hodnotení, pričom polovica bola tvorená negatívnymi hodnoteniami (hodnotenia z intervalu 0 až 1) a polovica pozitívnymi hodnoteniami (hodnotenia rovné 5). Uvedené hodnoty boli pri diskretizácii hodnôt pôvodnej vzorky zvolené na základe predpokladu jednoznačnej orientácie hodnotenia pri týchto krajných hodnotách. Vzorka bola tvorená hodnoteniami z rôznych oblastí, pričom presné zastúpenie jednotlivých tematických domén nebolo zisťované.Na základe analýzy niekoľkých vybraných hodnotení sme však overili, že vzorka obsahovala aj hodnotenia filmov.

Ako testovacia vzorka údajov bola použitávzorka údajov z práce Pang a Lee[9]. Ide o vzorku hodnotení zloženú z 1000 pozitívnych a 1000 negatívnych hodnotení filmov. Táto vzorka údajov je jednou z najčastejšie používaných vzoriek pri porovnávaní metód analýzy sentimentu. Presnosť klasifikácie dosiahnutá v práci [9] bola 82,9. Hoci jednotlivé filmy je možné žánrovo ďalej kategorizovať, z pohľadu ďalších experimentov bola vzorka považovaná za monotematickú vzorku, ktorej téma bola označená ako film.

[1]http://www.reviewcentre.com/

B. Postup

V prípade prvej metódy nebola zohľadňovaná tematická podobnosť hodnotení trénovacej a testovacej vzorky a overovaný bol iba vplyv veľkosti vzorky na presnosť klasifikátora. Z hodnotení bolo vytvorených 9 náhodných vzoriek vo veľkostiach 1% až 9% a 10 náhodných vzoriek vo veľkostiach 10% až 100% pôvodnej vzorky. Vytvorené klasifikátory boli aplikované na všetkých 2000 hodnotení z testovacej vzorky.

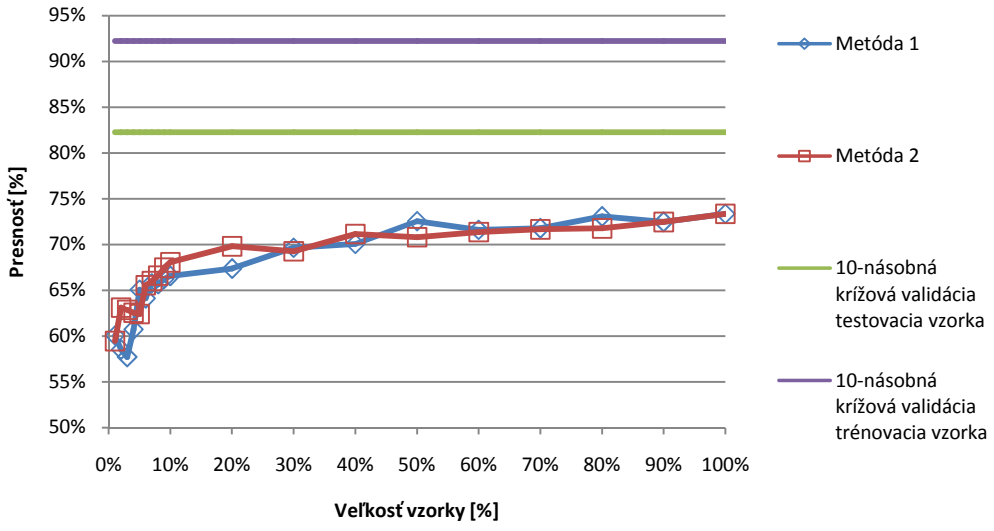
V prípade druhej metódy bolo na hodnotenia v testovacej vzorke okrem štandardnej tokenizácie a odvodenia koreňa slova (Stemming) aplikované aj značkovanie slovných druhov (Part-Of-SpeechTagging). Následne sme extrahovali iba podstatné mená s dĺžkou minimálne 3 znaky. Získané podstatné mená boli zoradené podľa frekvencie ich výskytu, ktorá bola znormalizovaná vzhľadom k celkovému počtu výskytov. Z výsledného zoznamu slov sme extrahovali prvých desať. Identifikované korene slov (film, movi, charact, atď.) veľmi úzko súviseli s tematickou doménou testovacej vzorky (film). Jednotlivé hodnotenia v rámci trénovacej vzorky boli následne ohodnotené na základe výskytu vážených slov, ktoré boli zvolené ako relevantné vzhľadom k identifikovanej doméne. Výhodou uvedeného riešenia je predovšetkým nízka výpočtová náročnosť. Zo zoradených hodnotení boli následne vytvorené rovnako veľké množiny slov ako v prípade prvej metódy.

Okrem presnosti porovnávaných metód boli určené aj presnosti pri použití 10-násobnej krížovej validácie aplikovanej nezávisle na trénovaciu a testovaciu vzorku. Tieto hodnoty predstavujúpriemerné presnosti dosiahnuté pri použití 9/10 pôvodnýchvzoriek na trénovanie a 1/10 na testovanie, pričom za testovaciu vzorku je použitá vždy iná desatina.

Pri všetkých experimentoch bola použitá metóda podporných vektorov. Použitá bola voľne dostupná implementácia LIBSVM typu *C-SupportVectorClassification* s lineárnou jadrovou funkciou, *epsilon* rovné 0.001 a*C* rovné 0. Ako metóda výberu atribútov bola použitá hodnota informačného zosilnenia.

V. VÝSLEDKY

Hodnoty presností, dosahované pri jednotlivých metódach sú zobrazené na obrázku 1.Horizontálna os zobrazuje percentuálnu veľkosť trénovacej vzorky údajov, pričom zobrazené sú hodnoty dosahované na intervale 0 až 0,1 a 0,1 až 1. Vertikálna os zobrazuje dosahované hodnoty presnosti. Z výsledkov je možné vidieť závislosť medzi veľkosťou trénovacej vzorky a dosahovanými hodnotami presnosti na testovacej vzorke. Táto závislosť je štatisticky významná na oboch intervaloch. Najvyššia hodnota presnosti (73,4%) bola dosiahnutá pri použití všetkých hodnotení v trénovacej vzorke. Na obrázku sú ako konštanty zobrazené aj priemerné presnosti dosiahnuté pri použití 10-násobnej krížovej validácie na trénovacej a testovacej vzorke. V prípade trénovacej vzorky bola dosiahnutá presnosť 92,22%, ktorá je prekvapivo vysoká a to predovšetkým vzhľadom k tematickej heterogenite vzorky. V prípade testovacej vzorky bola dosiahnutá presnosť 82,25%, ktorá je v súlade s výsledkami predošlých prác.



Obr. 1 Závislosť medzi veľkosťou trénovacej vzorky a presnosťou klasifikátora.

Predpokladaný nárast presnosti pri výbere hodnotení na základe ich tematickej podobnosti s testovacou vzorkou nebol štatisticky významným.Malé rozdiely v hodnotách presností je možné vysvetľovať z pohľadu použitej vzorky, ako aj použitej metódy pre výpočet podobnosti. Pôvodná vzorka nemusela obsahovať dostatočné množstvo tematicky relevantných hodnotení, hoci stránky Reviewcenter obsahujú aj hodnotenia filmov. Na druhej strane mohla byť nevhodná aj použitá metóda hodnotenia tematickej relevantnosti hodnotení. Zaradenie hodnotení do kategórií produktov nebolo pri vyhodnotení známe a preto nie je možné jednoznačne určiť správnosť metódy.

VI. ZÁVER

Automatická tvorba korpusov na základe on-line hodnotení sa na zvolenej vzorke testovacích údajov ukázala ako vhodné riešenie. Výsledky potvrdili zvyšovanie presnosti klasifikátorov v závislosti na veľkosti takto vytvorených korpusov. Keďže najvyššia presnosť klasifikátorov bola dosiahnutá pri najvyššom počte hodnotení, z výsledkov nie je možné určiť hranice zvyšovania presnosti. Jedným z cieľov ďalšieho výskumu by preto malo byť určenie tejto závislosti na základe analýz realizovaných na väčšom počte on-line hodnotení. Zvýšenie presnosti pri výbere trénovacej vzorky hodnotení na základe tematickej podobnosti voči témam hodnotení v testovacej vzorke sa neukázalo ako štatisticky významné. V rámci ďalšieho výskumu by sme preto chceli extrahovať témy hodnotení zo stránok a zároveň tak umožniť optimalizáciu metód pre výpočet tematickej podobnosti hodnotení. Uvedené experimenty by navyše bolo vhodné aplikovať na viacerých z voľne dostupných testovacích korpusov.

POĎAKOVANIE

Táto práca bola podporovaná Vedeckou grantovou agentúrou Ministerstva školstva, vedy, výskumu a športu Slovenskej republiky a Slovenskej akadémie vied na základe zmluvy č. 1/1147/12 (40%), Agentúrou na podporu výskumu a vývoja na základe zmluvy č. APVV-0208-10 (30%) a realizáciou projektu Rozvoj Centra informačných a komunikačných technológií pre znalostné systémy (ITMS kód: 26220120030) na základe podpory operačného programu Výskum a vývoj financovaného z Európskeho fondu regionálneho rozvoja (30%).

LITERATÚRA

[1] P. Koncz, "Podpora rozhodovania s využitím aspektovo založenej analýzy sentimentu," presented at the 6 th Workshop on Intelligent and Knowledge Oriented Technologies (WIKT 2011), Herľany, 2011.

[2] P. Koncz and J. Paralič, "Automatická analýza sentimentu textových zdrojov," presented at the Electrical Engineering and Informatics 2 : proceeding of the Faculty of Electrical Engineering and Informatics of the Technical University of Košice, Košice, Slovakia, 2011.

[3] M. Hu and B. Liu, "Mining and summarizing customer reviews," presented at the Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, Seattle, WA, USA, 2004.

[4] Q. Miao, *et al.*, "AMAZING: A sentiment mining and retrieval system," *Expert Systems with Applications*, pp. 7192-7198, 2009.

[5] J. Kamps, *et al.*, "Using WordNet to Measure Semantic Orientations of Adjectives," in *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC 2004)*, Lisbon, Portugal, 2004, pp. 1115-1118.

[6] S.-M. Kim and E. Hovy, "Automatic Detection of Opinion Bearing Words and Sentences," in *Proceedings of the Second International Joint Conference on Natural Language Processing (JCNLP 2005)*, Jeju Island, Korea, 2005, pp. 61-66.

[7] P. Koncz and J. Paralic, "An approach to feature selection for sentiment analysis," presented at the 15th IEEE International Conference on Intelligent Engineering Systems (INES 2011) Poprad, Slovakia, 2011.

[8] S. Tan and J. Zhang, "An empirical study of sentiment analysis for chinese documents," *Expert Systems with Applications*, vol. 34, pp. 2622-2629, 2008.

[9] B. Pang and L. Lee, "A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts," presented at the Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics, Barcelona, Spain, 2004.

Znalostný sklad pre predmet objavovanie znalostí

Peter KONCZ, Ján PARALIČ

Katedra kybernetiky a umelej inteligencie, Fakulta elektrotechniky a informatiky,
Technická univerzita v Košiciach, Slovenská republika

peter.koncz@tuke.sk, jan.paralic@tuke.sk

***Abstrakt* — Systémy pre správu obsahu predstavujú častý spôsob publikácie informácií a svoje uplatnenie nachádzajú aj ako stránky predmetov. Tieto systémy však primárne nie sú určené pre podporu celej škály procesov realizovaných v procese výučby. Využitie týchto systémov pre podporu vzdelávacích procesov si preto vyžaduje úpravu existujúcich a tvorbu nových rozšírení. Cieľom popisovaného projektu mine::YourFuture, podporeného Nadáciou Tatra banky v rámci grantového programu Kvalita vzdelávania, je vytvorenie a naplnenie znalostného skladu pre predmet objavovanie znalostí. Ako základ znalostného skladu je použitý voľne dostupný systém pre správu obsahu Joomla. Hlavná časť realizovaných prác súvisí s tvorbou samotného znalostného skladu, jeho obsahu v podobe video-tutoriálov, ako aj komponentov pre podporu ostatných procesov spojených s výučbou v rámci cvičení. Cieľom článku je popísať doposiaľ realizovanéa plánované práce, pričom osobitná pozornosť je venovaná znalostnému skladu a navrhovanému formálnemu modelu reprezentácie znalostí a anotácií v ňom.**

***Kľúčové slová* — znalostné modely, fazetová klasifikácia, systémy pre správu obsahu, manažment znalostí.**

I. ÚVOD

Cieľom projektu mine::YourFuture je vybudovanie verejne dostupného znalostného skladu pre podporu štúdia objavovania znalostí a práce s nástrojmi pre objavovanie znalostí. Hlavnou formou šírenia znalostí budú krátke video-tutoriály, priamo nadväzujúce na procesy popisované v prednáškach pozvaných odborníkov z oblasti objavovania znalostí. Používateľom tak bude poskytnutý jednak globálny pohľad na proces objavovania znalostí v podobe organizovaných prednášok odborníkov z praxe a jednak detailné popisy jednotlivých krokov a ich aplikácie v konkrétnom nástroji v podobe video-tutoriálov. Na pridávaní znalostí do znalostného skladu budú svojimi zadaniami participovať aj študenti. Vytvorený znalostný sklad by pritom mal slúžiť nielen študentom vybraných študijných odborov, ale aj výskumným pracovníkom univerzity, či zástupcom súkromného sektora, ktorí budú mať možnosť špecifikovať vlastné výskumné problémy a riešiť ich v spolupráci s pracovníkmi univerzity a študentmi. V rámci technologickej implementácie by zvláštna pozornosť mala byť venovaná vývoju znalostného skladu. Súčasťou projektu by okrem množiny študijných materiálov a technologickej platformy mal byť aj model procesu pre tvorbu, monitorovanie a udržiavanie znalostného skladu. Vytvorené riešenie by pritom malo umožňovať jeho prenositeľnosť na iné predmety. V práci sú najskôr popisované jednotlivé komponenty pre podporu procesov v rámci cvičení a následne je bližšie popisovaný komponent znalostného skladu.

II. REALIZÁCIA PROJEKTU

Vyššie formulovaný cieľ projektu je realizovaný v niekoľkých čiastkových cieľoch. Hlavným z nich je realizácia predmetového portálu spojeného s verejne dostupným znalostným skladom. Za základ portálu bol použitý opensourceCMS (ContentManagementSystem)Joomla[1]. Joomla je jedným z najpopulárnejších systémov tohto druhu, disponuje veľkým množstvom rozšírení a poskytuje tiež MVC (Model ViewControler) framework pre podporu tvorby vlastných komponentov. V rámci úvodných fáz projektu boli overené možnosti použitia samotnej platformy a vyskúšané viaceré z jej voľne dostupných rozšírení.

Aktuálne sú vytvárané komponenty, poskytujúce podporu procesov v rámci výučby na cvičeniach, ktorú sa nepodarilo zabezpečiť existujúcimi voľne dostupnými rozšíreniami.

[1]http://www.joomla.org/

Medzi uvedené funkcionality patríspráva skupín študentov, podpora prepojenia s LMS (LearningManagementSystem) Moodle, správa zadaní, správa dochádzky, ako aj tvorba reportov. Komponent správy skupín študentov je čiastočne nahraditeľnývlastným systémom Joomla pre správu používateľov, je však dôležitý z pohľadu funkcionality ostatných komponentov. Jeho funkcia spočíva vo vytváraní jednotlivých skupín študentov, ku ktorým budú následne vzťahované údaje z ostatných komponentov. V súčasnosti sú testy k predmetu realizované v LMS Moodle, ktorý by mal byť s vytváraným systémom prepojený. V aktuálne realizovanej podobe pôjde o jednoduché prepojenie na úrovni databáz. Identifikácia prostredníctvom univerzitného LDAP servera umožňuje automatizáciu prepojenia záznamov študentov v oboch systémoch. Komponent zadania v aktuálnej podobe podporuje definovanie typov zadaní, ich inštancií, ako aj priradení zadaní a výber zadania. Komponent dochádzka by mal umožňovať definovanie jednotlivých udalostí a následne sledovanie prítomnosti študenta na udalosti či už manuálnym zakliknutím vyučujúceho, alebo prostredníctvom zadania kľúča študentmi. Napokon komponent pre tvorbu reportov by mal automatizovať tvorbu reportov súvisiacich s dochádzkou a prospechom študentov.

Ďalším z cieľov projektu je tvorba videí prezentujúcich jednak prípadové štúdie použitia objavovania znalostí v praxi, ako aj jednotlivé kroky procesu objavovania znalostí. Prípadové štúdie budú prezentované v rámci pozvaných prednášok odborníkov z oblasti objavovania znalostí, ktorí popíšu proces aplikácie objavovania znalostí na nimi realizovaných prípadoch. Táto forma prednášok je v rámci predmetu realizovaná už dlhšie, v rámci projektu by však tieto prezentácie mali byť aj zaznamenanéa sprístupnené v podobe videí. Študentom tak bude umožnené získať globálny pohľad na proces objavovania znalostí. Jednotlivé kroky procesu objavovania znalostí zas budú priblížené v samostatných video-tutoriáloch v rozsahu 5 až 10 min. Doposiaľ vytvorené video-tutoriályprezentujú prácu v nástrojiRapidMiner, no pribudnúť by mali tutoriály pre SPSS Clementine.

Na pridávaní materiálov do znalostného skladu sa okrem vyučujúcich predmetu podieľajú aj študenti. Ich zadania sú totiž publikované ako články pozostávajúce z popisu jednotlivých metód využívaných v procese objavovania znalostí. Popisy metódsú tvorené teoretickým popisom metódy, príkladmi jej praktického použitia, ako aj exportom procesov ilustrujúcich ich použitie vo zvolenom nástroji. Touto formou študenti odovzdávali zadania už aj v priebehu zimného semestra a na základe skúseností s vytvorenými zadaniami sa zvolené CMS ukázalo ako vhodné pre uvedený účel.

Vytvorený portál by nemal slúžiť iba študentom predmetu, ale mal by tiež poskytovať verejne prístupnú kolaboratívnu a komunikačnú platformu pre transfer znalostí v oblasti objavovania znalostí. Prostriedkom zvýšenia záujmu študentov o predmet by mala byť aj pripravovaná súťaž v objavovaní znalostí, realizovaná prostredníctvom vytvoreného portálu. V záujme zvyšovania kvality výučby predmetu sa študenti budú môcť na konci semestra zúčastniť anonymného hlasovania, v rámci ktorého budú môcť poskytnúť spätnú väzbu k poskytovaným službám.

A. Znalostný sklad

Komponent pre správu znalostného skladu je jedným z najdôležitejšíchv rámci realizovaného projektu.Samotná Joomla disponuje iba možnosťou anotácie kľúčovými slovami, ako aj klasifikácie dokumentov do jednotlivých kategórií.Uvedené možnosti sme sa preto rozhodli rozšíriť vytvorením vlastného komponentu. Z pohľadu používateľa pridávajúceho znalosti by vytvorený komponent mal umožňovať:

- Tvorbu jednoduchých znalostných modelov.
- Anotáciu zdrojov vzhľadom k vytvorenému znalostnému modelu.

Z pohľadu požívateľa získavajúceho znalosti zo znalostného skladu by komponent mal umožňovať:

- Fulltextové vyhľadávanie.
- Prehľadávanie prostredníctvom mračna tagov.
- Prehľadávanie riadené hierarchiou konceptov.
- Atribútovo riadené vyhľadávanie/prehľadávanie.

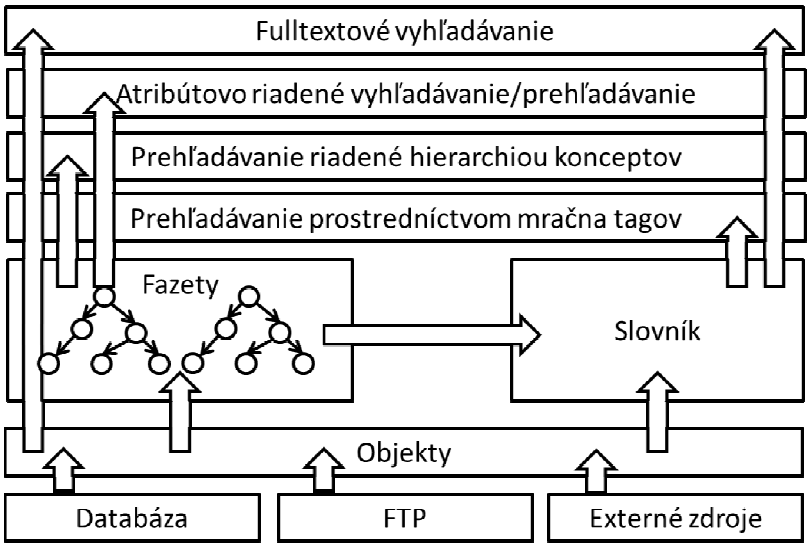
Prvou úlohou bola preto voľba formalizmu pre vyjadrenie znalostného modelu, ktorý by bol intuitívny a jednoducho použiteľný a zároveň by si zachoval výrazovú silu potrebnú vzhľadom k cieľom projektu. Zvolený formálny model pre reprezentáciu znalostného modelu a anotáciívychádza z prác Koncz[1] a Sacco a kol. [2] a je znázornený na obrázku 1.

Objekty. Pod objektmi sú v navrhovanom modeli chápané akékoľvek objekty určené k anotácii. V rámci komponentu Joomla budú objekty predstavované údajovými objektmi v databáze (články a pod.) na FTP serveri (videá, obrázky a pod.) ako aj externými zdrojmi.

Slovník. Slovník je tvorený množinou všetkých slov použitých v rámci anotácie objektov. Môže byť doplnený o údaje ako početnosť výskytu či väzba na jednotlivých anotátorov. Hoci slová implicitne odkazujú na koncepty, nie sú konceptmi (viď nižšie) a môžu slúžiť k označeniu

viacerých významovo rozdielnych konceptov. Jednotlivé slová nemusia byť priradené ku konceptom. Toto riešenie umožňuje použitie jednoduchých kľúčových slov v rámci anotácie.

Fazety. Zvolenou formou organizácie informácií sú v rámci navrhovaného komponentu fazety. Fazety je možné definovať ako „jasne definované, vzájomne sa vylučujúce a súhrne vyčerpávajúce aspekty, vlastnosti, alebo charakteristiky triedy alebo špecifického subjektu“ [3, s. 394]. V navrhovanom riešení používatelia nie sú ohraničovaní žiadnym metadátovým štandardom. Jednotlivé fazety predstavujú hierarchie konceptov, zodpovedajúce špecifikám organizácie informácií v rámci danej inštalácie. Požiadavka organizácie informačných zdrojov s ohľadom na špecifickú charakteristiku je realizovaná vytvorením príslušnej fazety. Každý z uzlov v rámci faziet je predstavovaný konceptom, ktorého význam je interpretovaný v kontexte danej fazety. Príkladom fazety v kontexte predmetu objavovanie znalostí je príslušnosť k niektorej fáze procesu objavovania znalostí, či príslušnosť ku konkrétnej hodine v rámci harmonogramu predmetu. Príkladom konceptu vo fazete „fáza procesu objavovania znalostí“ je „predspracovanie“.



Obr. 1 Navrhovaný model reprezentácie znalostí a anotácií a jeho prepojenie na znalostný sklad.

III. ZÁVER

V práci sú prezentované čiastočné výsledky prebiehajúceho projektu mine::YourFuture, zameraného na vytvorenie a naplnenie znalostného skladu pre predmet objavovanie znalostí. Popisované sú doposiaľ realizované ako aj plánované práce. Spustenie riadnej prevádzky portálu je plánované na začiatok zimného semestra nasledujúceho akademického roka na adrese mining.fei.tuke.sk. Vytvorené komponenty by po ukončení vývoja mali byť voľne dostupné na oficiálnej stránke rozšírení Joomla, ako aj na uvedenej projektovej stránke.

POĎAKOVANIE

Táto práca vznikla v rámci projektu mine::YourFuture, podporeného Nadáciou Tatra banka v rámci grantového programu Kvalita vzdelávania.

LITERATÚRA

[1] P. Koncz and K. Furdík, "Kategorizačný systém pre podporu kolektívnej anotácie internetových zdrojov," presentedattheElectricalEngineering and Informatics : ProceedingoftheFacultyofElectricalEngineering and InformaticsoftheTechnicalUniversityof Košice Košice, Slovakia, 2010.

[2] G. M. Sacco and Y. E. Tzitzikas, *DynamicTaxonomies and FacetedSearch. Theory, Practice, and Experience*vol. 25: Springer, 2009.

[3] A. G. Taylor and D. P. Miller,*Introduction to cataloging and classification*. Westport: LibrariesUnlimited, 2006.

Prístupy k vnímaniu emócií v humanoidnej robotike

Peter KONCZ

Katedra kybernetiky a umelej inteligencie, Fakulta elektrotechniky a informatiky,
Technická univerzita v Košiciach, Slovenská republika

peter.koncz@tuke.sk

***Abstrakt* — Algoritmická implementácia vnímania, modelovania a vyjadrovania emócií sa v poslednom období stáva čoraz diskutovanejšou, na čo reaguje aj táto práca. Práca je venovaná problematike vnímaniu emócií v oblasti humanoidnej robotiky a všeobecne v oblasti HumanComputerInteraction (HCI). Jej cieľom je poskytnúť stručný prehľad teoretických modelov emócií, používaných v uvedenej doméne, ako aj spôsobov ich implementácie pri rozpoznávaní emócií. Zároveň má popísať praktické využívanie resp. potenciálne možnosti využitia uvedených spôsobov a ďalších metód pre prácu s emóciami v rámci humanoidnej robotiky.**

***Kľúčové slová* — humanoidná robotika, rozpoznávanie emócií, humancomputerinteraction.**

I. ÚVOD

Podľa niektorých štúdií [1] je úplne neemocionálnych menej ako 10% ľudského života. Veľká časť emócií pritom nie explicitne vyjadrená, prejavujú sa však vo výraze tváre, intenzite a orientácii pohľadu, gestách, či hlase [2, 3, 4]. Väčšina týchto prejavov je vrodených a spoločných pre ľudský rod. Kroskulturálnosť výrazu emócií potvrdili okrem iného aj Ekman (tu podľa [2]). Jeho Dôkaz spočíval v experimentoch, v ktorých praobyvatelia Novej Guiney dokázali správne rozpoznať emocionálne výrazy na fotografiách príslušníkov iných, nimi nepoznaných kultúr, a naopak americký študenti dokázali rozoznať emocionálne výrazy týchto domorodcov. Popularita aplikácií využívajúcich emócie pri interakcii s používateľom pritom stále narastá, čo platí aj pre vývoj robotov [3, 5]. Táto skutočnosť je zrejme dôsledkom prenikania robotov do nových oblastí života ako sú domácnosti, zábava, zdravotná starostlivosť a pod. [4]. Praktické aplikácie, popisované v nasledujúcom texte, pritom často vychádzajú z rozličných teórií emócií, čomu zodpovedá aj spôsob delenia resp. merania charakteristík emócií pri ich vnímaní a rozpoznávaní. V nasledujúcej časti práce sú uvedené základné teórie emócií, používané v oblasti HCI a humanoidnej robotiky. Následne uvádzam základné techniky pre rozpoznávanie emócií, rozdelené podľa prejavu (výraz tváre, reč, gestá) v ktorom sú rozpoznávané. V tejto časti tiež uvediem niektoré zo súborov údajov, používaných pri trénovaní a hodnotení presnosti postupov pre rozpoznávanie emócií. Napokon uvádzam prehľad oblastí, v ktorých sú tieto techniky použiteľné resp. už používané.

II. TEÓRIE EMÓCIÍ

V súčasnosti neexistuje všeobecne prijímaná teória emócií, ale viaceré navzájom často konkurujúce teórie [2, 6]. Xiao a kol. [6] navyše uvádzajú, že v oblasti rozpoznávania emócii v reči, sa zvykne obyčajne konštatovať problém s klasifikáciou emócií, avšak zvolené členenie emócií je napokon často aplikačne závislé. Toto konštatovanie však platí aj pre ostatné oblasti rozpoznávania emócií.

Napriek nejednotnosti teoretických prístupov k emóciám, je však možné vyčleniť dve základné skupiny prístupov [7], ktorými sú diskrétne teórie a dimenzionálne teórie.

- Podľa Scherera[7] počiatky *diskrétnych teórií* je možné hľadať v práci Darwina, na ktoré nadväzovali práce Tomkinsa, Ekmana a Izarda. Spoločným pre túto skupinu teórií je, že rozlišujú konečný počet (medzi 9 až 14) základných emócií, ktorým je možné priradiť špecifické fyziologické, výrazové, či hlasové charakteristiky. Džuka[2] v tejto súvislosti hovorí o rozlišovaní obyčajne 6 – 10 bazálnych emócií, ktorými sú napríklad v prípade Ekmana hnev, odpor, strach, radosť, smútok a prekvapenie.
- Druhú skupinu predstavujú *dimenzionálne teórie*, ktorých počiatky siahajú k práci Wundta[7]. Ich spoločným znakom je vyčlenenie základných dvoch až troch dimenzii emócii, ktorými sú valencia, aktivita, prípadne ako tretia dimenzia sa zvykne uvádzať sila

resp. kontrola[6, 7]. Džuka[2] v tejto súvislosti hovorí o troch dimenziách v prípade systému navrhovanom Wundtom (príjemný / nepríjemný, vzrušenie / uspokojenie a napätie / uvoľnenie) a dvoch v prípade systému navrhovanom Bottenbergom (príjemný / nepríjemný, aktívny / pasívny).

- Podľa niektorých autorov [2, 3, 6] je možné ako tretiu primárnu skupinu teórií emócií, vyčleniť teórie založené na komponentoch (componentialtheories / models). Tieto teórie, na rozdiel od modelov založených na dimenziách, nie sú obmedzované počtom dimenzii a zároveň umožňujú vnútornú diferenciáciu v rámci jednej emócie resp. skupiny emócií[6], uznávajú teda význam tak kategórií ako aj dimenzií [3]. Mayring (podľa Džuka,2003) vychádza z päťkomponentovej štruktúry (situácie vyvolávajúce príslušnú emóciu, typické kognície, prežívanie, fyziologické koreláty, emocionálny výraz), kde komponenty nie sú izolované a sú chápané iba ako východisko pri lepšom popise a pochopení emócií.

III. TECHNIKY ROZPOZNÁVANIA EMÓCIÍ

Jedným z prístupov k deleniu techník pre rozpoznávanie emócií, je ich delenie na základe prejavu, v ktorom sú emócie rozpoznávané. Toto kritérium bolo použité aj v rámci tejto práce.

A. Rozpoznávanie emócií z výrazu tváre

Podľa [8], existujú tri hlavné triedy prístupov pre rozpoznávanie výrazov tváre:

- *Techniky založené na pohybe obrazu* (imagemotion). Ide o analýzu smeru a intenzity pohybu jednotlivých obrazových bodov na tvári.
- *Techniky založené na anotomických modeloch* (anatomicalmodels). Tieto techniky využívajú anatomické modely tváre, na interpretáciu výrazu (používané aj pri syntéze). Pri tejto skupine techník je pritom náročnou predovšetkým tvorba robustného modelu, valídneho pre široké spektrum možných tvárí.
- *Techniky založené na umelých neurónových sieťach* (neuralnetworks). Ako naznačuje názov tejto skupiny techník, ide o techniky využívajúce umelé neurónové siete.
- *Hybridné techniky* predstavujúce kombináciu predošlých.

Lisetti a Schiano [8] tieto techniky ďalej rozdeľujú podľa dvoch hlavných teoretických smerovaní na:

- *Konekcionistické prístupy*. Zastúpené sú technikami využívajúcimi umelé neurónové siete. Pri spracovaní obrazu sa zvyčajne vychádza z reprezentácie prostredníctvom obrázka (image-basedrepresentation).
- *Nekonekcionistické prístupy*. Zastúpené sú predovšetkým prostredníctvom techník, založených na pohybe obrazu a anatomických modeloch. Pri spracovaní obrazu sa vychádza z reprezentácie prostredníctvom vlastností (feature-basedrepresentation).

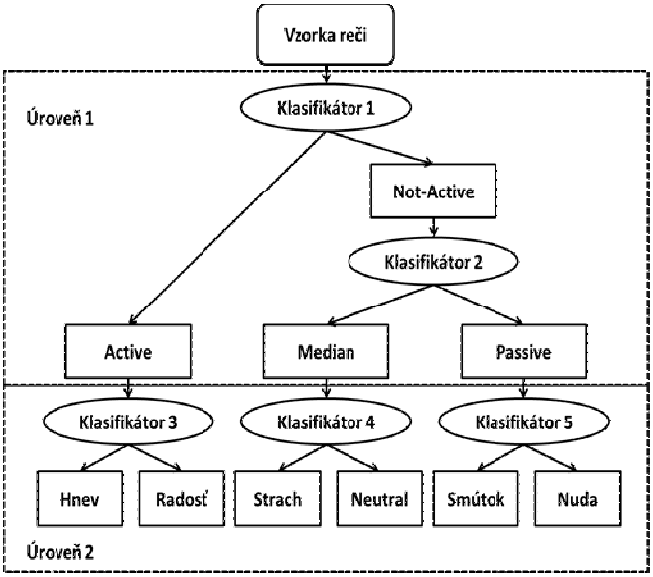
Mnohé z robotických systémov, využívajúcich výraz „tváre“ pri vyjadrovaní emócií, používa systém FACS (FacialActionCodingSystem)[3, 9], ktorý bol pôvodne vyvinutý Ekmanom a Friesenom. Tento systém je však možné využívať aj pri vnímaní emócií.

B. Rozpoznávanie emócií z vlastností reči

Vplyv emócií na charakteristiky hovorenej reči je zrejmý už stáročia, hoci o empirickom výskume je možné hovoriť až od začiatku 20. storočia, kedy sa psychiatri pokúšali o diagnostiku emocionálnych porúch, na základe elektroakustickej analýzy reči[7]. Podľa Fong a kol.[3], sú primárnymi parametrami, regulujúcimi emocionálny obsah v rámci hovoreného prejavu, hlasitosť a výška tónu (úroveň, variácia, rozsah) a prozódia. Murraya a Arnotta (tu podľa [3]) uvádzajú, že prejav emócií v hovorenom slove je konzistentný medzi rozličnými ľuďmi a sú medzi nimi iba malé rozdiely.

Polzina a Waibela[10] vo svojej práci popisujú návrh systému pre rozpoznávanie emócii v ľudskej reči, pričom analýza signálu je realizovaná v troch rovinách: na úrovni slov, prozodických vlastností reči a na úrovni frekvenčnej analýzy (spektrum). Systém klasifikuje tri typy emócií (neutrálna, smútok, hnev) s presnosťou, ktorá je podľa autorov porovnateľná s presnosťou človeka. Vo svojej práci následne popisujú implementáciu navrhovaného systému v dialógovom agentovi (eventuálne aj robot – poznámka autora), ktorý následne prispôsobuje diskusiu vnímanému emočnému kontextu.

Xiao a kol. [6]vo svojej práci popisujú návrh systému pre rozpoznávanie emócií, vychádzajúci z dimenzionálneho prístupu ku rozlišovaniu emócií. Na rozdiel od klasických riešení, využívajúcich pre klasifikáciu predovšetkým frekvenciu a silu signálu, ktoré sú významné predovšetkým z pohľadu dimenzie aktivácie, je autormi navrhovaný model doplnený o využívanie harmonických funkcií a tzv. Zipf funkcií. Ďalšia inovácia navrhovaného riešenia spočíva podľa autorov vo využívaní viacúrovňovej klasifikačnej schémy (obrázok 1). Klasifikačná schéma má na jednotlivých úrovniach binárne klasifikátory, využívajúce čo najvhodnejšie diskriminujúce funkcie.

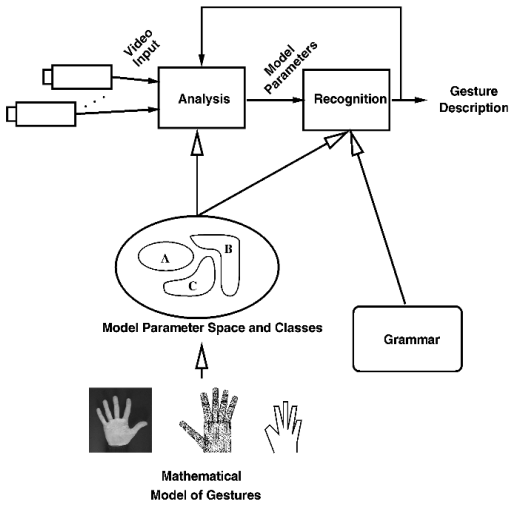


Obr. 1 Dimenzionálny emočný klasifikátor, pre klasifikáciu vzorov z BerlinEmotionalSpeechDatabase (podľa [6]).

Rozpoznávanie emócií v hovorenej reči, je možné kombinovať s rozpoznávaním emócií na základe iných charakteristík pozorovaného objektu, napríklad jeho výrazu tváre. Uvedené tvrdenie je samozrejme platné aj pri rozpoznávaní emócií na základe iných prejavov.
Príkladom praktického využitia rozpoznávania emócii v hovorenej reči je aj call-centrum, ktoré dokáže presmerovať nahnevaného zákazníka priamo na operátora[6].

C. Rozpoznávanie emócií z gest

Pod gestami (gestures), sú v oblasti HCI myslené „dynamické a/alebo statické konfigurácie ľudskej ruky, ramena a tiež iných častí ľudského tela, ktoré sú merateľné strojmi“[11]. Nedostatkom tejto definície však je, že vylučuje robota ako gestikulujúceho.
Pavlovic a kol.[11], rozdeľujú postupy používané pri rozpoznávaní gest na dve skupiny. Prvú skupinu predstavujú postupy, ktoré priamo merajú (pomocou snímačov umiestnených na snímanej osobe) priestorové a temporálne charakteristiky rúk a ramien. Nedostatkom tohto prístupu je predovšetkým obmedzovanie používateľa, ktorý je nútený pre interakciu s počítačom resp. robotom nosiť prístroje, obyčajne rukavice, z čoho vyplýva aj autormi použité označenie tejto skupiny prístupov, ako prístupov reprezentovaných tzv. *na rukaviciach založenými zariadeniami* (glove-baseddevices). Za vhodnejšiu preto považujem druhú skupinu techník, ktoré autori označili ako „na videu založené nekontaktné interakčné techniky“ (video-basednoncontactinteractiontechniques). Autori, vychádzajúc z existujúcich výskumov, uvádzajú základnú všeobecnú štruktúru systému pre rozpoznávanie gest (viď obrázok 2), ktorá je použiteľná aj pre oblasť rozpoznávanie emócií (rozšírenie modelu o ďalšie časti, využívajúce gestá pri analýze emócií).



Obr. 2 štruktúra systému pre rozpoznávanie gest pomocou videozáznamu (podľa [11]).

Prvou etapou pri návrhu systému je *vytvorenie modelov gest*, kde sa využívajú zväčša matematické modely. Tieto modely zachytávajú tak priestorové, ako aj dynamické charakteristiky gest, hoci ako uvádzajú Pavlovic a kol.[11], v čase ich výskumu dominoval statický popis. Tvorba modelu pritom spravidla predchádza samotnému pozorovaniu gest, resp. emócií na základe gest. V rámci *analýzy gest* pozorovaného objektu (človek, robot, atď.), sa potom určia parametre, zodpovedajúce použitému modelu. Následne dochádza k *rozpoznávaniu gest* na základe parametrov, akceptovaného modelu, ako aj prijatých gramatických pravidiel. Pod *gramatickými pravidlami* je myslená tak vnútorná syntax gest, ako aj popis interakcie gest s inými komunikačnými spôsobmi, ako je reč, pohľad, výraz tváre a pod. Pre detailnejší popis jednotlivých fáz rozpoznávania resp. subsystémov systémov pre rozpoznávanie pohybu viď[11].

D. Rozpoznávanie emócii na základe fyziologických ukazovateľov

Vstupom do týchto systémov sú rôzne fyziologické reakcie, ako napr. kožná odporová reakcia (SCR – Skin ConductanceResponse), záznam elektrokardiogramu (ECG), záznam elektroencefalogramu (EEG), srdcová frekvencia, krvný tlak a iné. Mnohé z fyziologických ukazovateľov súvisiacich s emóciami, je pritom možné merať iba resp. prednostne prostredníctvom priameho kontaktu senzorov s telom sledovaného subjektu. Navyše cena prístrojov pre snímanie niektorých vstupných premenných, ako napr. EEG či ECG, je relatívne vysoká. Tieto skutočnosti sú zrejme hlavnými príčinami malého počtu prác, ktoré pre rozpoznávanie emócií využívajú fyziologické charakteristiky. Vo výskume Westerink a kol.[12] boli ako fyziologické vstupy použité údaje zo zápästného snímača kožnej odporovej vodivosti a pásového snímača srdcovej frekvencie a výstupom boli predikované emočné stavy. Tieto stavy boli porovnávané so stavmi uvádzanými účastníkmi výskumu, ktorý ich zadávali pomocou prenosného zariadenia.

IV. SÚBORY TRÉNOVACÍCH A TESTOVACÍCH ÚDAJOV

S otázkou rozpoznávania emócií a voľby modelu ich členenia je úzko spätá otázka súborov údajov, používaných pri trénovaní klasifikátorov a prediktorov aplikovaných pri rozpoznávaní emócií resp. hodnotení ich výkonnosti (tzv. benchmarkingové súbory údajov). V prípade rozpoznávania emócií ide, v závislosti od skúmaného indikátora emócií, napr. o zvukové záznamy reči pri rozpoznávaní emócií v reči, obrázky resp. video záznamy pri rozpoznávaní emócií z výrazov tváre, posturality či gest, alebo o záznamy EEG či kožnej vodivosti a pod. pri rozpoznávaní emócií na základe fyziologických prejavov. Anotácie záznamov potom obsahujú údaje o príslušnosti záznamu k niektorej z emócií (diskrétny prístup), či sýtení jednotlivých dimenzii emócií (dimenzionálny prístup), prípadne ďalšie informácie.

Berlínska databáza emocionálnej reči (BerlinEmotionalSpeechDatabase) je tvorená záznamami desiatich nemeckých viet z bežnej komunikácie, ktoré sú piatimi hercami a piatimi herečkami nahovorené v siedmich emocionálnych verziách[13]. Autormi databázy je použitý prístup, využívajúci súbor diskrétnych emócií (neutrálna, hnev, strach, radosť, smútok, odpor, nuda). Databáza je voľne dostupná na adrese http://www.expressive-speech.net/emodb.

Dánska databáza emocionálnej reči (DanishEmotionalSpeechDatabase) je tvorená záznamami slov, viet a pasáží v Dánštine, ktoré sú dvomi hercami a dvoma herečkami nahovorené v piatich emocionálnych verziách[14]. Použitý bol súbor diskrétnych emócií (neutrálna, prekvapenie, radosť, smútok, hnev). Databáza je pre nekomerčné účely dostupná na adrese http://cpk.auc.dk/~tb/speech/Emotions/.

Tieto súbory údajov je možné podľa spôsobu získania záznamov rozdeliť do nasledovných kategórií[7]: záznamy získané z prirodzenej interakcie, záznamy indukovaných emócii a v súčasnosti najčastejšie sa vyskytujúce záznamy simulované profesionálnymi resp. amatérskymi hercami. Tieto kategórie uvádza Scherer[7] v súvislosti s emocionálnymi zvukovými súbormi údajov, avšak toto delenie je platné aj pre iné typy emocionálnych databáz (obrazových, filmových a pod.). Rozsiahly prehľad desiatok zvukových emocionálnych databáz, s informáciami o jazyku záznamov, nahrávaných subjektoch (herci, pacienti, deti a pod.), doplňujúcich signáloch (EEG, galvanická odpoveď kože, krvný tlak, srdcový tep a pod.), účeloch záznamu (rozpoznávanie, syntéza) a spôsobe získania emócií (viď vyššie), je možné nájsť v práci Ververidisa a Kotropoulosa[15]. Tento prehľad teda zahŕňa aj databázy, obsahujúce údaje použiteľné pri rozpoznávaní emócií na základe fyziologických ukazovateľov.

V oblasti rozpoznávania emócií z výrazu tváre neexistuje databáza, ktorá by bola jednotne prijímaná odborníkmi z danej oblasti[16]. Databázou používajúcou už zmienený kódovací systém FACT je databáza FACS Dictionary. Medzi ďalšie patria JapaneseFemaleFacialExpression (JAFFE) database, JapaneseCaucasianFacialExpressionsofEmotions (JACFEE) database, PicturesofFacialAffect (POFA) database, či Cohn-Kanadefacialexpressiondatabase[17].

V. PRAKTICKÉ APLIKÁCIE ROZPOZNÁVANIA EMÓCIÍ

Nižšie popisované praktické aplikácie nevyužívajú iba techniky používané pri rozpoznávaní emócií. Tieto aplikácie zväčša kombinujú techniky určené k rozpoznávaniu emócií, ich vnútornej reprezentácii resp. modelovaniu, ako aj vyjadrovaniu. Popisované praktické aplikácie je možné rozdeliť do nasledujúcich oblastí:

Regulácia správania a sebaochrana

Prejavmi emócií v prípade človeka sú okrem iného regulácia intenzity a orientácie pozornosti, ako aj rozhodovacích procesov a niektorých vrodených behaviorálnych prejavov[2, 18]. Dôležitou funkciou týchto mechanizmov vplyvu emócií na človeka je vlastná ochrana organizmu. Picard[18] v tejto súvislosti uvádza prácu CynthieBreazeal, ktorá pracuje na návrhu riadenia filmového robota. Tento robot disponuje 61 stupňami voľnosti a v dôsledku toho mu v niektorých prípadoch hrozí poškodenie spôsobené vlastnými pohybmi. Cieľom je teda implementácia takých regulačných mechanizmov, podobných emóciám, ktoré robotovi zabránia vo vykonaní nebezpečných pohybov. Podobný účel emócií je možné pozorovať aj v prípade robota, popisovaného Canamerom[5]. Cieľom tohto robota, simulovaného vo virtuálnom prostredí, bolo čo najdlhšie „prežiť“ a snažiť sa udržať stabilitu svojho vnútorného prostredia (hladinu krvného cukru, cievnu mohutnosť, energiu a pod.). Jeho správanie bolo pritom regulované súborom na prežitie orientovaných motívov a základných emócií.

Reinforcementlearning

Canamero vo svojej práci kladie otázku „ ako môže robot interpretovať vnímaný signál z okolia sám a nepoužívať (resp. nebyť odkázaný na – poznámka autora) odmeňujúcu informáciu poskytovanú nejakým typom vonkajšieho učiteľa“[5]. Za možné riešenie považuje používanie simulovaných emócií.

Pamäť

Ďalšou oblasťou, v ktorej môže byť výhodné využitie umelých emócií, je pamäť robotov, ktorá pri nevhodnej forme reprezentácie informácií a ich nedostatočnom filtrovaní môže byť preťažená. V tejto oblasti môžu byť inšpiráciou modely vplyvu emócií na pamäť u človeka[5].

Výskum emócii

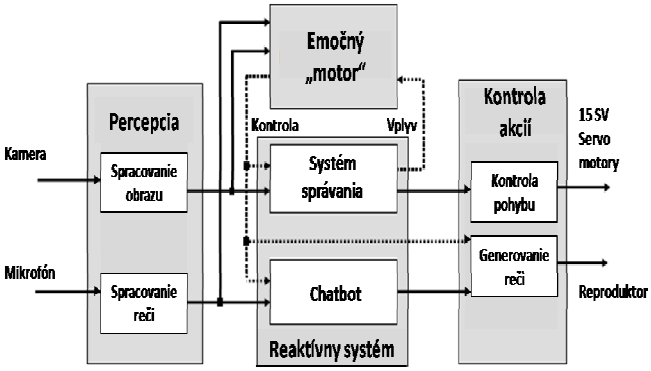
Vzťah podpory medzi výskumom emócii a implementáciou emócii v robotike je recipročný, pričom podporovanými oblasťami výskumu emócií môžu byť[5]:

- Ľudské vnímanie emócií. Ide o využitie expresívnych robotov, ktorí dokážu upútať človeka a vyvolať u neho emocionálnu odpoveď. Aj veľmi jednoduché nástroje pritom stačia na to, aby mohlo byť skúmané ľudské vnímanie emócií a ich význam pre ľudskú tendenciu k antropomorfizácii (Reeves aNass, 1996 tu podľa [5]).
- Nástroje pre testovanie a prešetrenie teórií. Roboty, v ktorých sú implementované umelé emócie, zväčša na základe nejakého teoretického modelu emócií, môžu slúžiť ako „virtuálne laboratóriá“. Simulácie s nimi by pomohli pri analýze teórie, na ktorej sú postavené. To môže byť zvlášť užitočné pokiaľ by reálne testy na reálnych subjektoch boli problematické (etické princípy, opakovateľnosť, izolovateľnosť a pod.).
- Syntetický prístup k emóciám. Ide o snahu o pochopenie emócií prostredníctvom ich syntézy. Je komplementárnym k analytickým prístupom.
- Operacionalizácia a zjednodušenie. Na rozdiel od čisto teoretických modelov je možné technicky realizované modely lepšie operacionalizovať a možno ich aj zjednodušiť, čo môže prispieť k vyzdvihnutiu podstatných aspektov teórie a jej lepšiemu pochopeniu.

Praktickými aplikáciami, ktoré je možné zaradiť do tejto skupiny, sú napríklad robotické tváre Kismet[4, 19, 20] a MEXI[21]. Kismet je emocionálne interagujúcou robotickou tvárou, vyvíjanou v rámci Massachusettského technologického inštitútu. Na základe audiovizuálnych vstupov dokáže rozpoznávať rôzne emocionálne podnety od ľudí vo svojom okolí a zároveň vysielať sociálne signály pomocou orientácie pohľadu, výrazu tváre, či prozódie[20]. Emočný systém Kismetu je pritom inšpirovaný viacerými modelmi emócií a motivácií ľudí aj zvierat[4, 19]. Na rozdiel od Kismetu, projekt MEXI nie je zameraný na simuláciu mechanizmov sociálnej interakcie ľudí a ich vývoja, ale sleduje skôr konštruktívny postoj, pričom cieľom je vytvorenie čisto reaktívneho správania na základe pohnútok a umelých emócií robota[21]. Vnímanie emócii je v prípade MEXI realizované analýzou prozódie reči, ako aj výrazu tváre (vo vývoji). Vyjadrovanie emócií je taktiež realizované prostredníctvom výrazu tváre a rečových prejavov. Schéma architektúry MEXI je znázornená na obrázku 3.

Starostlivosť o ľudí, rehabilitácia a terapia

Je ďalšou perspektívnou oblasťou pre využívanie emocionálne interagujúcich robotov. Príkladom takéhoto robota je rehabilitačný robot Robbie, popisovaný v práci Mohana a kol.[22]. Do tejto kategórie aplikácií je možné zaradiť aj robota mediátora, ktorý učí autistické detí základné interakčné zručnosti [23].



Obr. 3 Schéma architektúry MEXI (podľa Esau, Kleinjohann, & Kleinjohann, 2006).

VI. ZÁVER

Strojmi realizované rozpoznávanie emócií, ich vnútorná reprezentácia, ako aj vyjadrovanie, má v oblasti humanoidnej robotiky nepochybne svoje miesto. Hoci, ako je to vidieť aj v tejto prehľadovej práci, výskum a vývoj v tejto oblasti sú iba na svojom počiatku. Dá sa však predpokladať, že postupné začleňovanie robotických systémov do každodenného života, bude výrazným motívom rozvoja tejto oblasti poznania.

LITERATÚRA

[1] R. Cowie. (2009, Emotion-orientedcomputing: state oftheart and keychallenges. *emotion-research.net*. Available: http://emotion-research.net/projects/humaine/aboutHUMAINE/HUMAINE%20white%20paper.pdf
[2] J. Džuka, *Motivácia a emócie človeka*. Prešov, 2003.
[3] T. Fong, *et al.*, "A surveyofsociallyinteractiverobots," *Robotics and AutonomousSystems*, pp. 143–166, 2003.
[4] C. Breazeal, "Emotion and sociablehumanoidrobots," *InternationalJournalofHuman-ComputerStudies*, pp. 119–155, 2003.
[5] L. Canamero, "Emotionunderstandingfromtheperspectiveofautonomousrobotsresearch," *NeuralNetworks*, pp. 445–455, 2005.
[6] Z. Xiao, *et al.*, "Recognitionofemotions in speech by a hierarchicalapproach," in *InternationalConference on AffectiveComputing and IntelligentInteraction*, 2009.
[7] K. R. Scherer, "Vocalcommunicationofemotion: A reviewofresearchparadigms," *SpeechCommunication*, pp. 227–256, 2003.
[8] C. Lisetti and D. Schiano, "Automaticfacialexpressioninterpretation: Wherehuman–computerinteraction, artificialintelligence, and cognitivescienceintersect," *Pragmatics and Cognition*, pp. 185-235, 2000.
[9] M. S. Bartlett, *et al.*, "AutomaticRecognitionofFacialActions in SpontaneousExpressions," *JournalofMultimedia*, pp. 22-35, 2006.
[10] T. S. Polzin and A. Waibel, "Emotion-sensitivehuman-computerinterfaces," *Proceedingsofthe ISCA workshop on Speech and Emotion*, pp. 201-206, 2000.
[11] V. I. Pavlovic, *et al.*, "VisualInterpretationofHandGesturesforHuman-ComputerInteraction: A Review," *IEEE transactions on pattern analysis and machineintelligence*, pp. 677-695, JULY 1997.
[12] J. Westerink, *et al.*, "Emotionmeasurementplatformfordailylifesituations," in *AffectiveComputing and IntelligentInteraction and Workshops, 2009*, 2009.
[13] F. Burkhardt, *et al.*, "A DatabaseofGermanEmotionalSpeech," in *ProceedingsInterspeech*, Lisbon,Portugal, 2005.
[14] I. S. Engberg and A. V. Hansen, *DocumentationoftheDanishEmotionalSpeechDatabase DES*. Aalborg, 1996.
[15] D. Ververidis and C. Kotropoulos, "Emotionalspeechrecognition: Resources, features, and methods," *SpeechCommunication*, pp. 1162–1181, 2006.
[16] M. Pantic and L. J. M. Rothkrantz, "FacialActionRecognitionforFacialExpression Analysis FromStatic Face Images," *IEEE transactions on systems, man, and cybernetics*, pp. 1449-1461, 6 2004.
[17] N. V. Afzulpurkar and A. R. Abbasii. (2009, HumanFacialExpression and EmotionRecognition. *ComputerScience and InformationManagement program*. Available: http://www.cs.ait.ac.th/~cs103343/Adv-ImageProcessing_Nitin.pdf
[18] R. W. Picard, "Affectivecomputing: challenges," *InternationalJournalofHuman-ComputerStudies*, pp. 55–64, 2003.
[19] C. Breazeal, "AffectiveInteractionbetweenHumans and Robots," *Advances in ArtificialLife*, pp. 582-591, 2001.
[20] C. Breasel, "Kismet," in *http://www.ai.mit.edu*, ed, 2009.
[21] N. Esau, *et al.*, "EmotionalCommunicationwiththe Robot Head MEXI," in *9th InternationalConference on Control, Automation, Robotics and Vision*, 2006
[22] R. E. Mohan, *et al.*, "EvaluatingVirtualEmotionalExpressionSystemsforHuman Robot Interaction in RehabilitationDomain," in *InternationalConference on Cyberworlds 2008*, 2008, pp. 554-560.
[23] I. Werry, *et al.*, "CanSocialInteractionSkillsBeTaught by a Social Agent? The Role of a RoboticMediator in AutismTherapy," in *ProceedingsoftheInternationalConference on CognitiveTechnology*, 2001.

Influence of learning parameter γ on Stochastic Update modifications of Backpropagation Through Time

Juraj KOŠČÁK, Rudolf JAKŠA, Peter Sinčák

Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

jurajkoscak@gmail.com, jaksa@neuron.tuke.sk, peter.sincak@tuke.sk

Abstract **– We will examine various modifications of backpropagation through time algorithm (BPTT) done by stochastic update in the artificial recurrent neural networks (ARCNN) and the influence of the different value of learning parameter γ. The general introduction involving the stochasticity into neural network was provided by Salvetti and Wilamowski in 1994 in order to improve probability of convergence and speed of convergence. The implementation is simple for arbitrary network topology. In stochastic update scenario, constant number of weights/neurons is randomly selected and updated. This is in contrast to classical ordered update, where always all weights/neurons are updated. Stochastic update is suitable to replace classical ordered update without any penalty on implementation complexity and with good chance without penalty on quality of convergence. We have provided first experiments with stochastic modification on backpropagation algorithm (BP) used for artificial feed-forward neural network (AFFNN) in detail described in the article [1] published on conference in Barcelona. We will present experiment results on simple toy-task data of postponed and modified signal as a verification of our implementation of different algorithm modifications.**

Keywords **– artificial recurrent neural networks (ARCNN), backpropagation algorithm (BP), backpropagation through time (BPTT), stochastic update.**

I. Introduction

The stochastic weight update was introduced by Salvetti and Wilamowski in 1994 [2] in order to improve probability of convergence and speed of convergence of the backpropagation algorithm. Besides the stochastic weight update, they examined another two stochastic methods: random pattern selection and randomized learning rate. On the XOR problem they demonstrated significant improvement in the learning speed and probability of convergence for every one from these methods, especially for randomized learning rate. The backpropagation algorithm (Werbos, 1974; Rumelhart, McClealland, 1986) is one of most used learning algorithms today. Plain vanilla implementation and momentum implementation (Silva, Almeida, 1990) [3] are the predominant implementations on feed-forward topologies. Multilayer perceptron (MLP) is most used feed-forward topology, it is a layered architecture with fully interconnected layers. The BPTT [4] is most used backpropagation variant for recurrent topologies. It is well suited algorithm for recurrent topologies with full connectivity. Although the BPTT implementation is not much more complex compare to feed-forward vanilla backpropagation, the feed-forward time-delay networks are preferably used for time-related problems. Time-delay networks are of the MLP design, so layered topology is most used topology with the backpropagation algorithm. Sparse topologies became recently popular with the recurrent echo state networks (ESN) (Jaeger, 2001) [5]. However these are usually not based on backpropagation algorithm. ESN networks are based on so called reservoir of randomly

II. Stochastic Weight Update

Salvetti and Wilamowski did incorporate stochastic processes into backpropagation algorithm in order to improve convergence - to avoid to get stuck in a local minima on error surface [2]. Stochastic Weight Update was one from several ways how to do it. During learning, maybe it is not necessary to update all weights. Not all of them have the same influence on the network error. Some of them are more important, maybe have higher values, some of them are close to

zero, maybe their influence is lower, or none. If we don't update all weights, we can also save some computation time and thus maybe speed up the learning process.

1) Ordered Update: Classic BPTT algorithm, where all weights are update in one iteration. Error signal is computed for all neurons without any preferences.
2) Stochastic Update: Modified BPTT algorithm, where not all weights are updated in the same iteration. The chance of weight update is controlled by r, where $0 \leq r \leq 100$, which represents the probability of weight update in percentage. The weight is selected for update if the generated number is $rand \leq r$. Otherwise the weight is not going to be updated. The ARCNN keeps this way information from the past.
3) Stochastic Δ Weight Update: This approach combines the Stochastic Update with the stochastic selection of δweights, that are going to update. After the back propagated error signal is computed, the δ weights are going to be update. The stochastic approach is the same as with Stochastic Update. δ weight is updated if the $rand \leq r$.
4) Stochastic δ Update: Combines Stochastic Update with stochastic selection of δ Update. Stochastic factor is the same, as in the rest of stochastic modifications. The δ is selected to be update with probability. This selective updating of δ cases, that also here some old δ from $t-1, ..., t-x$ (x - not known the past, that haven't been update yet) acts in the learning process and have influence on the whole process.
5) Stochastic Complex Update: The stochastic factor of probability is the same as in previous stochastic methods. It combines Stochastic Update, Stochastic δWeight Update and Stochastic δ Update.
6) Shuffle δ Update: This approach simplifies the implementation BPTT algorithm. It is more simple because you do not have to care about layers and the order of the neurons. Simply one neuron is randomly selected from the "pool of neurons", error signal is computed and all incoming weights are updated. The "pool of neurons" represents all neurons in all layers, that contribute on δ calculation.

If we want to increase weight selection chance over 100% (all weights will be updated) we can add another loop with parameter, which control how many times this random selection should be done. If set to 1 it is equal to the loop missing.

III. Experiments

To provide the basic experiments of Stochastic Weight Update implementation, we chose the modified and postponed signal, a toy-task benchmark see Fig. 1, that proves the correct implementation of the ARCNN. We want to examine, that the Stochastic Update modifications, will not reduce quality of learning, and also to explore the influence of different γ value of weights updated per pattern on the results. The results of Stochastic update are compared to the origin Ordered Update implementation of BPTT algorithm on ARCNN. We use the same number of neurons, the same sigmoid function with the steepness λ. The learning parameter γ is change experiment by experiment from $0.03, 0.1, 0.3, 0.5, 0.7\ and\ 0.9$, together six experiment. Figure 2 till Figure 7 represents the graphical results of the experiment and detailed settings of ARCNNs are in the Table 1. Figure 8 till Figure 13 represents the Least mean square error (LMS) which is used as the comparison between different γ value experiments. Lower LMS provides better result.

IV. Conclusion

The conclusion of the experiment is, that the Stochastic Update modifications in different variants is less sensitive to the different learning parameter γ compare to classical ordered update. Still there is some influence of the γ which depends on the implementation level of stochasticity. This can be used as an advantage during the tuning of ARCNN and we can be more focused on the topology itself and data structure used for of the ARCNN. To find out the suitable γ value can be postponed till we have more or less known working topology. In some cases, this less dependence of ARCNN on the γ can theoretically improve accuracy of results produced by ARCNN.

Complexity of stochastic update implementation is less dependent on the network topology than the ordered update is, while overall complexity of implementation is not much higher than the classical one. This can be utilized in BPTT implementations with non-standard, sparse, or random network topologies.

Our main goal is to provide the prediction of the temperature daily profile using the implemented BPTT stochastic update, mainly the Shuffle Update, due to the ability of ARCNN to use the history of previous steps stored inside. We already prepared the testing dataset for temperature prediction, that consist of the measured temperature from the year 2000 till the end of February 2011. The

data are measured each half hour, which provides us wide range of options to create suitable training and testing data for ARCNN. This wide range also allows the ARCNN to create internal model of daily temperature profile.

REFERENCES

[1] J. Koščák, R. Jakša, and P. Sinčák, "Stochastic weight update in the backpropagation algorithm on feed-forward neural networks," in *Neural Networks (IJCNN), The 2010 International Joint Conference on*, 2010.

[2] A. Salvetti and B. M. Wilamowski, "Introducing stochastic processes within the backpropagation algorithm for improved convergence," in *ANNIE'94 - Artificial Neural Networks in Engineering, (Intelligent Engineering Systems Through Artificial Neural Networks)*, C. Dagli, B. Fernández, J. Ghosh, and R. Kumara, Eds., vol. 4. ASME PRESS, New York, 1994, pp. 205–209.

[3] F. Silva and L. Almeida, "Acceleration technique for the backpropagation algorithm," in *Neural Networks EURASIP workshop*, 1990.

[4] P. Werbos, *The Roots of Backpropagation: From Ordered Derivation To Neural Networks and political forecasting*, February 1994.

[5] H. Jaeger, "Gmd report 148, the "echo state" approach to analysing and training recurrent neural networks," GMD - German National Research Institute for Computer Science, Tech. Rep., 2001.

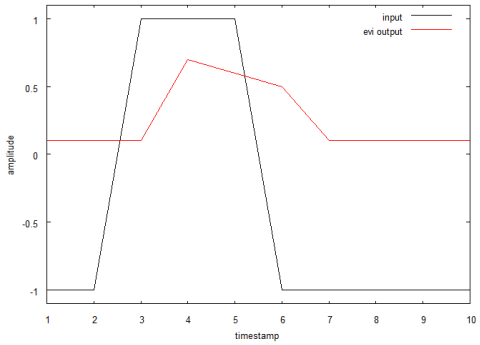


Fig. 1 Moved and modified signal used for learning ARCNN.

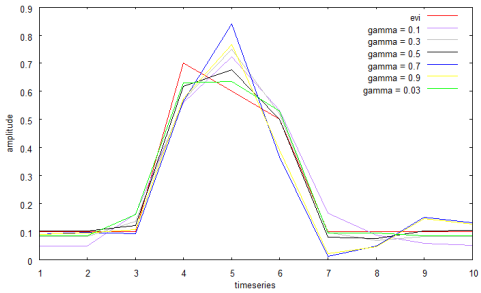


Fig. 2 The influence of the learning parameter γ on the learning process using Ordered Update.

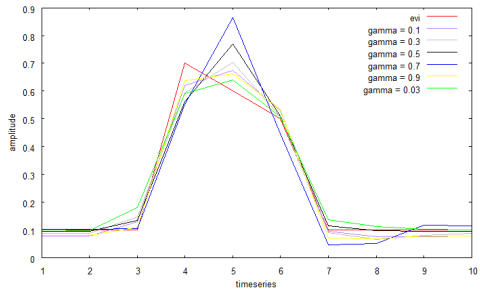


Fig. 3 The influence of the learning parameter γ on the learning process using Stochastic Update.

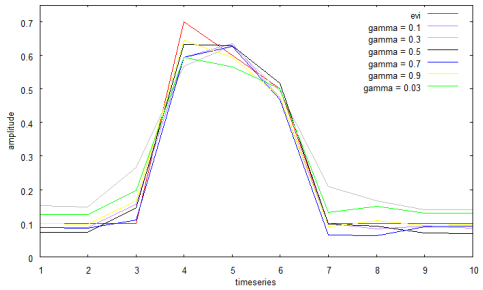


Fig. 4 The influence of the learning parameter γ on the learning process using Stochastic ΔWeight Update.

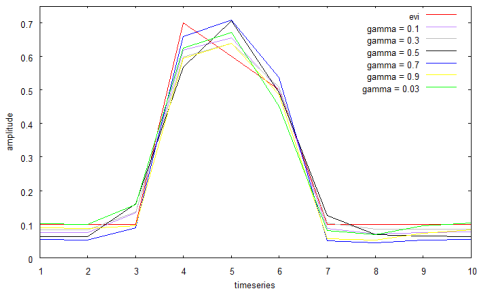


Fig. 5 The influence of the learning parameter γ on the learning process using Stochastic Complex Update.

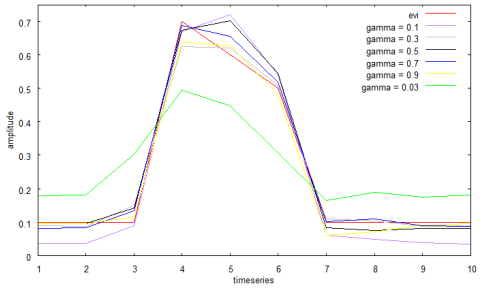


Fig. 6 The influence of the learning parameter γ on the learning process using Stochastic δ Update.

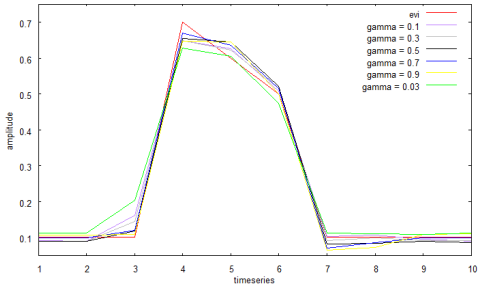


Fig. 7 The influence of the learning parameter γ on the learning process using Shuffle δ Update.

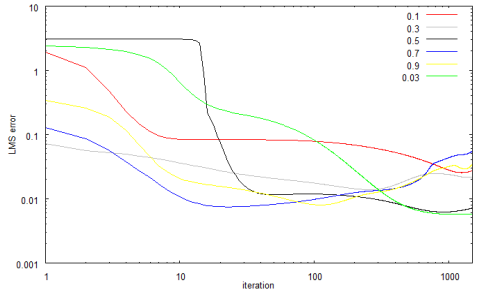


Fig. 8 LMS error of the influence of the learning parameter γ on the learning process using Ordered Update.

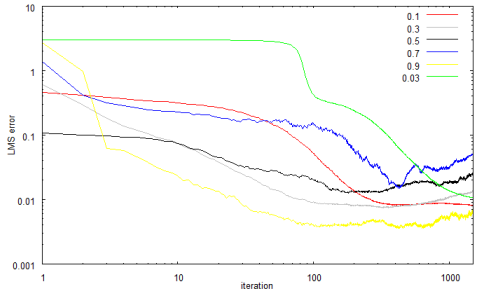


Fig. 9 LMS error of the influence of the learning parameter γ on the learning process using Stochastic Update.

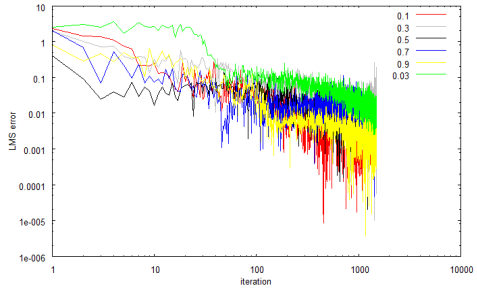


Fig. 10 LMS error of the influence of the learning parameter γ on the learning process using Stochastic ΔWeight Update.

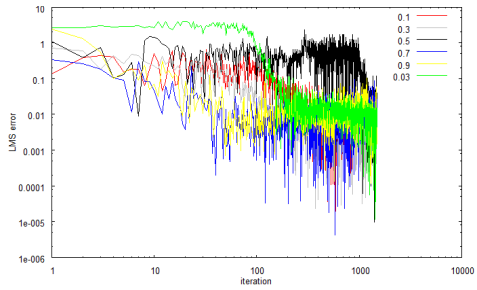


Fig. 11 LMS error of the influence of the learning parameter γ on the learning process using Stochastic Complex Update.

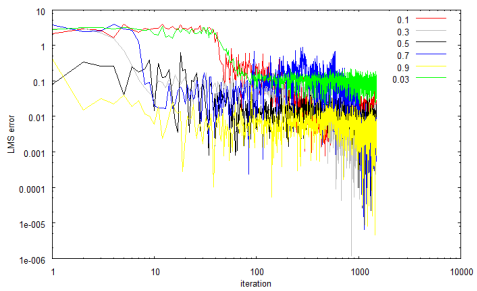


Fig. 12 LMS error of the influence of the learning parameter γ on the learning process using Stochastic δ Update.

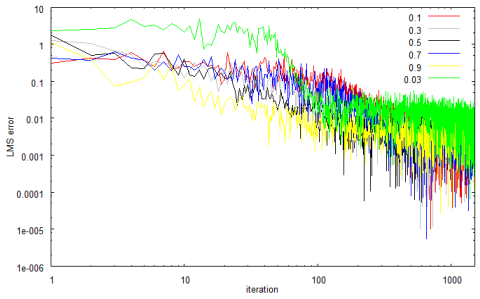


Fig. 13 LMS error of the influence of the learning parameter γ on the learning process using Shuffle δ Update.

Table 1
Influence of γ on Stochastic Update modifications

Method	Ordered Update	Stochastic Update	Stochastic ΔWeight Update	Stochastic Complex Update	Stochastic δUpdate	Shuffle δUpdate
Selection Ratio	100%	50%	50%	50%	50%	50%
γ 1	0.1	0.1	0.1	0.1	0.1	0.1
γ 2	0.3	0.3	0.3	0.3	0.3	0.3
γ 3	0.5	0.5	0.5	0.5	0.5	0.5
γ 4	0.7	0.7	0.7	0.7	0.7	0.7
γ 5	0.9	0.9	0.9	0.9	0.9	0.9
λ	0.65	0.65	0.65	0.65	0.65	0.65
α momentum	0.0	0.0	0.0	0.0	0.0	0.0
Unfolding history T	1	1	1	1	1	1
Topology	1-15-1	1-15-1	1-15-1	1-15-1	1-15-1	1-15-1
Adaptaptation in one cycle	1	1	1	1	1	1
Iterations	1500	1500	1500	1500	1500	1500
LMS error 1	0,0271	0,0081	0,0082	0,012	0,017	0,0020
LMS error 2	0,0216	0,013	0,0310	0,0073	0,0035	0,0033
LMS error 3	0,0071	0,0247	0,0052	0,0086	0,017	0,0027
LMS error 4	0,0542	0,0504	0,0027	0,0212	0,0016	0,0030
LMS error 5	0,0341	0,0063	0,0044	0,0017	0,0022	0,0066

The influence of the learning parameter γ on the learning process. The stochastic update modifications are less sensitive to the γ, than the ordered update.

Návrh a realizácia informačného systém pre flexibilný montážny podnik

Tomáš KRAJČI, Matej ČOPÍK, Ján JADLOVSKÝ

Katedra kybernetiky a umelej inteligencie, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

tomas.krajci@student.tuke.sk, matej.copik@tuke.sk, jan.jadlovsky@tuke.sk

***Abstrakt* — Tento článok sa venuje návrhu a realizácii informačného systém pre flexibilný montážny podnik. Informačný systém komunikuje s databázou Oracle 10g pomocou JDBC ovládača od firmy Oracle a s PLC automatom pomocou JEasyOpc implementácie. Súčasťou práce je vytvorenie informačného systému, ktorý umožňuje spracovávať objednávky produktov a odosielať výrobné príkazy priamo výrobnej linke. Aplikácia je navrhnutá a realizovaná vo vývojovom prostredí Eclipse v programovacom jazyku JAVA s podporou frameworkov Spring, Hibernate a Quartz.**

***Kľúčové slová* — MIS, JAVA, Spring, PLC, Oracle, Quartz, časovač, web**

I. ÚVOD

Programovací jazyk JAVA čoraz viac nachádza zastúpenie v tvorbe informačných systémov na báze enterprise architektúry. Vzhľadom na schopnosti tohto programovacieho jazyka ovládať aj zariadenia bola táto voľba logická a keďže som úspešne dokázal komunikovať s PLC automatom pri tvorbe svojej bakalárskej práce, vybral som si opäť programovací jazyk JAVA.

II. POPIS FLEXIBILNÉHO MONTÁŽNEHO PODNIKU

FMP tvorí výrobná linka, ktorá zostavuje produkty a ukladá ich do skladu.

A. Produkt

Produkt FMP pozostáva zo štvorcovej hliníkovej podstavy s otvorom, do ktorého je zasadené ložisko. V stredovom otvore tohto ložiska je zasunutá nylonová hriadeľka na ktorej je zasadený klobúčik. Každý z komponentov produktu je vyrábaný v dvoch verziách. Hrubé a tenké ložiská, zelené a modré hriadeľky a hliníkové alebo železné klobúčiky. O tomto výrobnom procese je viac v diplomovej práci Bc. Igora Jadlovského [4].

B. Výroba

Sériová výroba FMP začína vsunutím podstavy do výrobného procesu. Táto podstava je potom uložená na jednu z paliet, ktoré sa presúvajú dookola na hlavných dvoch dopravníkových pásoch. Po vložení postavy na paletu sa paleta presunie na ďalší post, kde je do nej zasadené ložisko. Následne sa paleta presunie na druhý dopravníkový pás a je zachytená ďalším postom, kde je do ložiska vložená hriadeľka. Na hriadeľku je potom nasunutý klobúčik. Celý tento produkt je nakoniec druhým dopravníkovým pásom dopravený k poslednému postu, kde je založený do skladu.

C. Informačný systém

Na informačnej úrovni FMP sa nachádza informačný systém, ktorý má za úlohu spracovávať objednávky, evidovať vyrobené produkty, chybové stavy linky a spravovať užívateľov.

ISBN 978-80-553-0890-6 © 2012 FEI TUKE

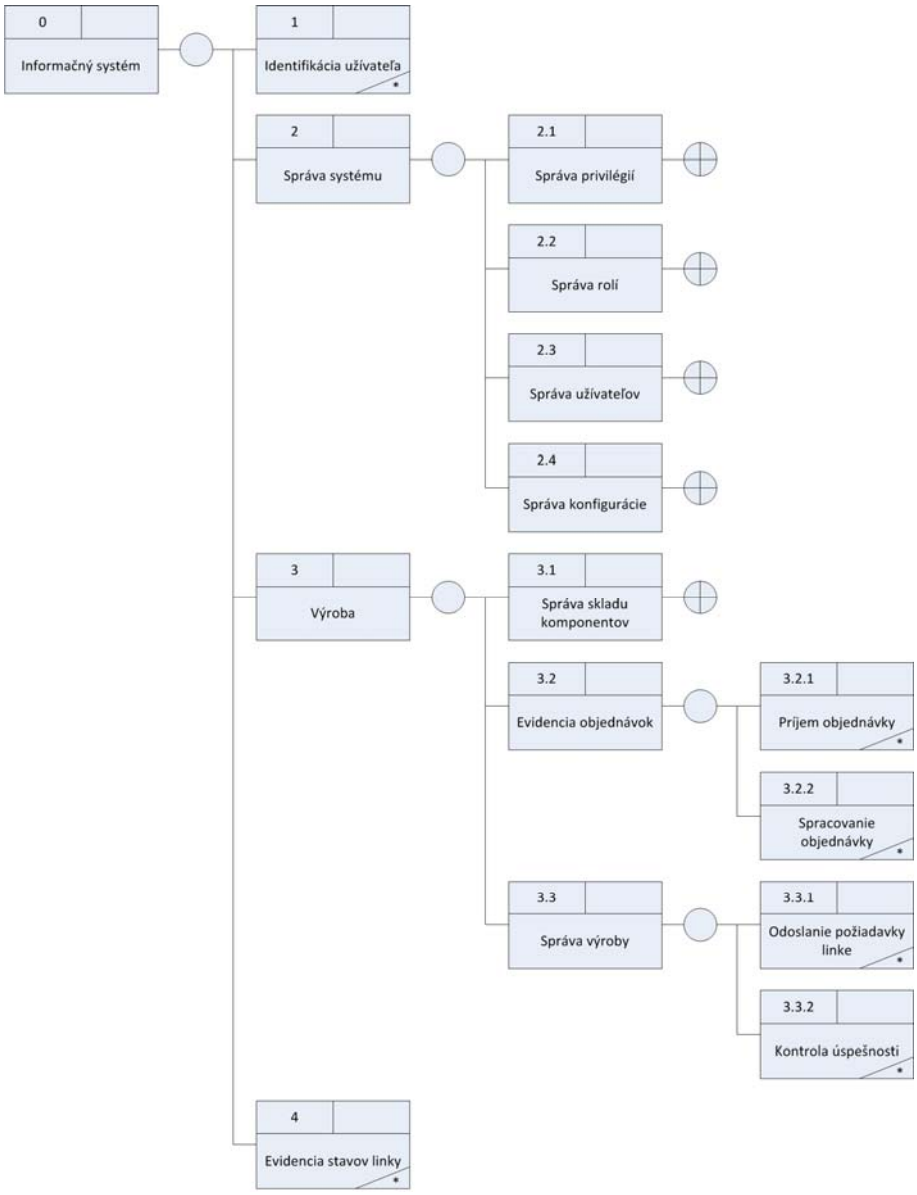
III. NÁVRH INFORMAČNÉHO SYSTÉMU

A. Dekompozícia procesov

Rozdelením resp. distribuovaním uceleného systému na menšie časti, ktoré nie sú ovládané centrálne ale lokálnymi riadiacimi systémami vytvárame distribuovaný systém riadenia. V dnešnej dobe z ekonomických dôvodov, ale aj z dôvodov spoľahlivosti je snaha vytvárať práve distribuované riadiace systémy. [1]

„Základným problémom decentralizovaných metód riadenia je dekompozícia. Zložitý systém je potrebné rozdeliť (dekomponovať) na N jednoduchších podsystémov.“ [2] Tieto podsystémy sú ovládané samostatnými procesormi, čo je prínosom z ekonomického hľadiska a tiež z hľadiska spoľahlivosti a efektivity. Existuje viacero spôsobov, ako tieto podsystémy medzi sebou komunikujú, napríklad pomocou ethernetovej siete. Tieto systémy majú tiež schopnosť pracovať nezávisle na sebe z toho dôvodu, aby pri výpadku jedného nevypadol celý systém. [3]

V diagrame (**Obr. 1**) je znázornená dekompozícia procesov informačného systému. Začína nultou úrovňou, ktorá znázorňuje celý informačný systém. Vzhľadom na užívateľa, prvý proces musí byť vždy identifikácia a autentifikácia. Vzhľadom na komplexnosť je systém rozdelený na viacero procesných vetiev a týmto spôsobom treba rozdeliť aj oprávnenia jednotlivých užívateľov.



Obr. 1: Diagram dekompozície procesov

Správa systému je ďalšou časťou systému. Tieto procesy sa týkajú hlavne fungovania systému, užívateľov a konfigurácie. Procesy správy privilégií sú pridávanie, modifikácia a mazanie. Správa rolí sa skladá z rovnakých procesov, ale pridáva sa aj proces prideľovania privilégií do danej role. Správa užívateľov sa taktiež skladá z procesov pridávania, modifikácie a mazania užívateľov a taktiež aj prideľovania rolí. Konfigurácia systému však obsahuje iba proces modifikácie nastavení, pretože konfiguračných volieb je dopredu stanovený počet.

Proces výroby obsahuje správu skladu komponentov. Sklad komponentov obsahuje komponenty používané vo výrobe, ako sú klobúčiky, hriadeľky a ložiská. Ďalej sa dekomponuje na procesy pridávania, zmeny a odoberania pre každý komponent a taktiež nastavenia množstiev vo vstupnom sklade. Správa objednávok obsahuje procesy objednávania a kontroly zodpovedným človekom za výrobu. Objednávateľ odošle objednávku a vždy ju musí potvrdiť vedúci výroby. Správa výroby sú procesy týkajúce sa samotnej výroby linkou. Obsahuje procesy odosielania požiadaviek linke a taktiež aj kontrolu úspešnosti výroby.

Evidencia stavov linky je proces, ktorý kontroluje stavy linky a eviduje ich v databáze.

B. Komunikačné rozhrania

Informačný systém komunikuje s užívateľom pomocou internetového prehliadača. Toto zjednodušuje jeho ovládanie, pretože na strane klienta netreba inštalovať žiaden ďalší software. S databázou Oracle 10g informačný systém komunikuje pomocou JDBC driveru od firmy Oracle a Hibernate frameworku. So samotnou linkou flexibilného montážneho podniku komunikuje IS cez rozhranie OPC pomocou implementácie JEasyOPC. Pomocou tejto implementácie sa IS napojí na RSLinx server, cez ktorý už má priamy prístup k externým premenným systému.

IV. REALIZÁCIE INFORMAČNÉHO SYSTÉMU

Informačný systém dodržuje trojúrovňovú enterprise architektúru. To znamená, že je rozdelený do troch vrstiev funkcionality, ktorými sú:

- DAO (Data Access Object) – slúži na prepojenie s databázou a PLC
- Manager – slúži na spravovanie internej logiky
- Web – reprezentačná vrstva
- Časovače – časovanie procesov

A. Entity

Entity sú vlastne objekty, s ktorými informačný systém pracuje. Väčšina entít systému je reprezentovaná tabuľkami v databáze. Tieto entity sú teda serializovateľné. Ostatné entity sú len dočasné a slúžia pre podporu programovej logiky. Serializovateľné entity sú rozdelené do troch kategórií podľa toho, ktorá časť systému ich využíva. Môže to byť užívateľský manažment, manažment objednávok alebo stav a konfigurácia systému.

Dočasné entity existujú iba v rámci jednej *session* alebo v rámci jednej úlohy systému, ako napríklad entita, ktorá obsahuje reťazec časovania.

B. DAO

Vo vrstve DAO sa nachádzajú triedy, ktoré sa starajú o získavanie dát z externých zdrojov. Či už ide o databázu alebo OPC konektivitu. V tomto informačnom systéme som použil dva typy DAO tried, ktorými sú databázové DAO a PLC DAO. Všetky DAO triedy majú vlastné rozhrania kvôli možnostiam prepájania vrstvy DAO s ostatnými vrstvami systému.

Skupina databázových DAO sa stará o prístup na Oracle databázu pomocou *Hibernate* frameworku. Tieto triedy obsahujú základné metódy prístupu, ako sú *persist*, *merge*, *find*, *findAll*, *remove*, *flush* a *refresh*. Trieda *AbstractDaoImpl* obsahuje všetky tieto základné implementácie, ktoré ostatné triedy iba dedia. Týmito metódami IS zaisťuje všetky potrebné úkony, ktoré systém potrebuje vykonávať nad databázou. Niektoré triedy DAO však potrebovali ešte špecifické prístupové metódy, ako napríklad vyhľadávanie podľa určitej premennej alebo limitovanie počtu výsledkov. Tieto metódy sa nachádzajú v každej jednotlivej DAO triede, ktorá ich potrebuje.

PlcDao je najkomplikovanejšie zo všetkých DAO objektov informačného systému. Nepracuje totiž s databázou ale s OPC protokolom, pomocou ktorého IS komunikuje s výrobnou linkou. Využíva implementáciu *JEasyOPC*, čo je knižnica, ktorá umožňuje

pripojenie na OPC server, v tomto prípade *RS Linx*. Toto DAO pristupuje na externé premenné linky, ktoré sú definované v rozhraní *PlcDao*. Táto DAO trieda obsahuje metódy inicializácie rozhrania ako aj metódy samotnej komunikácie s PLC automatom. Tieto metódy kontrolujú stav externých premenných automatu a taktiež umožňujú im nastavovať nové hodnoty.

C. Managers

Vrstva Managers je podstatná v tom, že spája všetky vstupno-výstupné vrstvy a obsahuje vnútornú programovú logiku informačného systému. Na tejto vrstve sa nachádzajú taktiež rozhrania a ich implementácie, pretože komponenty tejto vrstvy tiež musia byť schopné naviazania na Web vrstvu. Manažéri na tejto vrstve sú rozdelené podľa ich rozsahu na *GUserManager*, *RoleManager*, *OrderManager*, *ProductManager*, *StatusManager* a *ConfigManager*.

Vrstva manažérov je dôležitá aj z hľadiska toho, že v určitých prípadoch treba modifikoať alebo inak pripraviť dáta pre prezentačnú vrstvu, ktoré dostávame z vrstvy DAO. Uplatnením tejto vrstvy teda máme centrálne miesto, kde dané úpravy prebiehajú.

D. Web

Prezentačnou vrstvou je v prípade tohto informačného systému vrstva Web. Táto vrstva používa technológiu Spring MVC (Model-view-controller). MVC umožňuje separáciu modelu, pohľadu a ovládania užívateľského vstupu. Táto architektúra umožňuje nezávislý vývoj, testovanie a údržbu každej z týchto častí. Vzhľadom na informačný systém to znamená, že časť je reprezentovaná priamo kódom v Jave a časť je reprezentovaná pomocou Java Server Pages (JSP). Vzhľadom na užívateľské rozhranie, som použil Apache Tiles 2 framework, ktorý mi umožnil rozdeliť JSP na časti a tie, ktoré sú platné pre celú Web vrstvu som mohol použiť viac krát.

E. Časovače

Časovače informačného systému sú realizované pomocou *Quartz* a *Spring Batch* frameworkov. *Spring Batch* spravuje *Job* a *JobRepository*, ktoré obsahujú úlohu odosielania objednávok výrobnej linke a celý proces výroby. *Quartz* sa stará o spúšťanie týchto úloh ako aj o pravidelnú kontrolu stavu linky. Konfigurácia týchto komponentov sa nachádza v XML súboroch popri ostatnej konfigurácii systému. Časovače sú vlastne na úrovni Managers, no vymykajú sa tým, že bežia nezávisle na interakciách užívateľa.

V. ZÁVER

V práci sa mi podarilo navrhnúť a realizovať informačný systém flexibilného montážneho podniku ako súčasť distribuovaného systému riadenia. Pri tvorbe som úspešne použil jazyk JAVA a tým som dokázal, že informačné systémy na JAVA platforme majú uplatnenie aj pri informačných systémoch, ktoré sú priamo napojené na výrobné linky. Informačný systém je popri štandardnej funkcionalite očakávanej od IS schopný aj inteligentne vyberať poradie odosielania požiadaviek s kombináciou rovnakých prototypov z viacerých objednávok súčasne. Triedenie prototypov však nespája len rovnaké prototypy, ale triedi ich aj na základe podobnosti, čiže najpodobnejšie za sebou nasledujú a to zjednodušuje pracovníkom dopĺňanie komponentov do vstupných zásobníkov.

POĎAKOVANIE

Táto práca bola vytvorená realizáciou projektu Rozvoj Centra informačných a komunikačných technológií pre znalostné systémy (kód ITMS projektu: 26220120030) na základe podpory operačného programu Výskum a vývoj financovaného z Európskeho fondu regionálneho rozvoja.

LITERATÚRA

[1] CHOVAŇÁK, J: Riadenie a vizualizácia modelu sústavy manipulátorov na dispečerskej a informačnej úrovni, Diplomová práca, KKUI FEI TU Košice, 2007
[2] SARNOVSKÝ, J.: Riadenie zložitých systémov, ALFA Bratislava, 1988.
[3] KRAJČI, T.: Viacuživateľský systém pre ovládanie a riadenie automatického regálového zakladača, Bakalárska práca, KKUI FEI TU Košice, 2010
[4] JADLOVSKÝ, Igor: Návrh a realizácia riadenia flexibilného montážneho podniku PLC automatom, Diplomová práca (ešte nepublikovaná), KKUI FEI TU Košice, 2012

Riešenie úloh s ohraničeniami ako problém splniteľnosti

Michal LECHMAN, Marián MACH

Katedra kybernetiky a umelej inteligencie, Fakulta elektrotechniky a informatiky,
Technická univerzita v Košiciach, Slovenská republika

michal.lechman@student.tuke.sk, marian.mach@tuke.sk

***Abstrakt* — Tento článok sa zaoberá riešením úloh s ohraničeniami prostredníctvom ich transformácie do tvaru úlohy splniteľnosti s následným riešením niektorým z existujúcich algoritmov pre riešenie SAT úloh. Podáva stručný teoretický prehľad problematiky s dôrazom na reprezentáciu úloh ako problémov splniteľnosti. Popisuje experimenty s rôznymi reprezentáciami testovacieho problému a na základe ich porovnania vyvodzuje odporúčania ohľadom reprezentácie ohraničení implicitným alebo explicitným spôsobom. Výsledkom tejto práce má byť poukázanie na dôležitosť samotnej reprezentácie problému pri jeho riešení.**

***Kľúčové slová* — problém s ohraničeniami, SAT, reprezentácia ohraničení, Sudoku**

I. ÚVOD

Pri riešení úloh s ohraničeniami je možno využiť pomerne veľký počet rozličných prístupov k vyhľadávaniu riešení úloh takéhoto typu [1]. Jedným z týchto existujúcich prístupov je transformačný prístup – problém s ohraničeniami je mapovaný na určitý štandardný model (štandardnú úlohu), tento model je následne vyriešený špecializovaným algoritmom a výsledky sú spätne transformované do pôvodnej oblasti . Úlohu takéhoto modelu môže hrať známy problém splniteľnosti (SAT). Pri takomto spôsobe riešenia do popredia vystupuje otázka vhodnej transformácie, ktorá nemusí byť vždy jednoznačná – výsledkom môže byť niekoľko rôznych reprezentácií pôvodného problému v SAT oblasti.

Cieľom tejto práce je skúmať dôležitosť reprezentácie problémov s ohraničeniami, dôraz je na skúmaní toho, ako daná reprezentácia vplýva na rýchlosť výpočtu.

Štruktúra článku je organizovaná nasledovne. V nasledujúcej kapitole je stručne definovaný SAT problém a rozoberaná otázka transformácie dvoch typov ohraničení do tvaru problému splniteľnosti. V ďalšej časti sa venujeme experimentovaniu s rôznymi reprezentáciami vybraného testovacieho problému vo vytvorenom testovacom prostredí.

II. SPLNITEĽNOSŤ A REPREZENTÁCIA OHRANIČENÍ

A. SAT (splniteľnosť booleovských formúl)

Veľké množstvo rozhodovacích problémov je možné transformovať na úlohu SAT, čiže na problém splniteľnosti booleovskej formuly. Ide o rozhodovací problém, ktorý patrí do oblasti výrokovej logiky – zaoberá sa vyhodnocovaním pravdivostných výrazov. Máme výraz, ktorý je zapísaný len pomocou operátorov reprezentujúcich konjunkciu, disjunkciu a negáciu, premenných a zátvoriek. Otázkou je či existuje k danému výrazu priradenie logických hodnôt jednotlivým premenným takým spôsobom, aby výraz bol ohodnotený ako pravdivý.

Zvyčajne sa používa vyjadrenie v tvare konjunktívnej normálnej formy (vo všeobecnosti vedúce na NP-úplný problém), ktorá spĺňa nasledovné podmienky:

- formula je zadaná ako konjunkcia klauzúl,
- každá klauzula reprezentuje dizjunkciu literálov,
- každý literál je buď premenná alebo jej negácia.

Pre splnenie formuly je potrebné dosiahnuť, aby v každej klauzule bola aspoň raz zastúpená pravdivá hodnota [2].

B. Reprezentácia ohraničení

V tomto článku sa zameriavame na ohraničenia typu „min 1“ a „max 1“. Každé z týchto ohraničení je možné definovať pre skupinu *n* premenných *V1*, ..., *Vn*. Ohraničenie typu „min 1“ vyžaduje, aby aspoň jedna z premenných *V1*, ..., *Vn* bola splnená. Takéto ohraničenie je možné reprezentovať jednou klauzulou (majúcou *n* členov) tvaru

V1 or ... or *Vn*

Ohraničenie typu „max 1“ stanovuje, že zo skupiny daných premenných môže byť splnená maximálne jedna premenná (alebo žiadna). Pre reprezentáciu takéhoto ohraničenia existuje niekoľko rôznych reprezentačných schém, napríklad binárna alebo sekvenčná reprezentačná schéma [3]. Najjednoduchšie kódovanie je binomické, ktoré nevyžaduje definíciu pomocných reprezentačných premenných. Pomocou tohto kódovania je možné ohraničenie vyjadriť ako sadu klauzúl tvaru

not(Vi) or *not(Vj)*

pre indexy *i=1,...,n-1* a *j=i+1,...,n*. Výsledkom je teda množina *n(n-1)/2* dvojčlenných klauzúl, kde každá z premenných je súčasťou *n-1* klauzúl.

III. EXPERIMENTY A VÝSLEDKY

A. Testovací problém

Ako testovací príklad bol použitý problém s ohraničeniami nad konečnými doménami, ktorý patrí do kategórie NP-úplných problémov - známy hlavolam sudoku. Tento hlavolam sa skladá z mriežky tvorenej deviatimi riadkami, deviatimi stĺpcami a deviatimi štvorcami, pričom štvorce majú rozmer 3x3. Hlavolam obsahuje 81 pozícií, kde pozícia 11 hovorí o ľavom hornom rohu a pozícia 99 zase o pravom dolnom rohu, pričom niektoré pozície sú obsadené nejakou číslicou, zatiaľ čo iné pozície sú ešte voľné. Cieľom tohto hlavolamu je pridať jednu číslicu od 1 do 9 na každú voľnú pozíciu tak, aby v žiadnom riadku, stĺpci ani štvorci neboli dve rovnaké číslice.

Pre naše účely je možné tento problém reprezentovať sadou premenných tvaru *Vxyz,* kde index *x* reprezentuje riadok, *y* stĺpec a *z* hodnotu umiestnenú na priesečníku daného riadku a stĺpca. Takýmto spôsobom máme k dispozícii 729 pravdivostných premenných. Ohraničenia je možné vyjadriť ako [4]:

a) na každej pozícii je prípustná maximálne jedna z možných hodnôt
b) na každej pozícii musí byť aspoň jedna z možných hodnôt
c) v každom riadku sa každá z hodnôt môže vyskytovať maximálne raz
d) v každom riadku sa každá z hodnôt musí vyskytovať aspoň raz
e) v každom stĺpci sa každá z hodnôt môže vyskytovať maximálne raz
f) v každom stĺpci sa každá z hodnôt musí vyskytovať aspoň raz
g) v každom štvorci sa každá z hodnôt môže vyskytovať maximálne raz
h) v každom štvorci sa každá z hodnôt musí vyskytovať aspoň raz

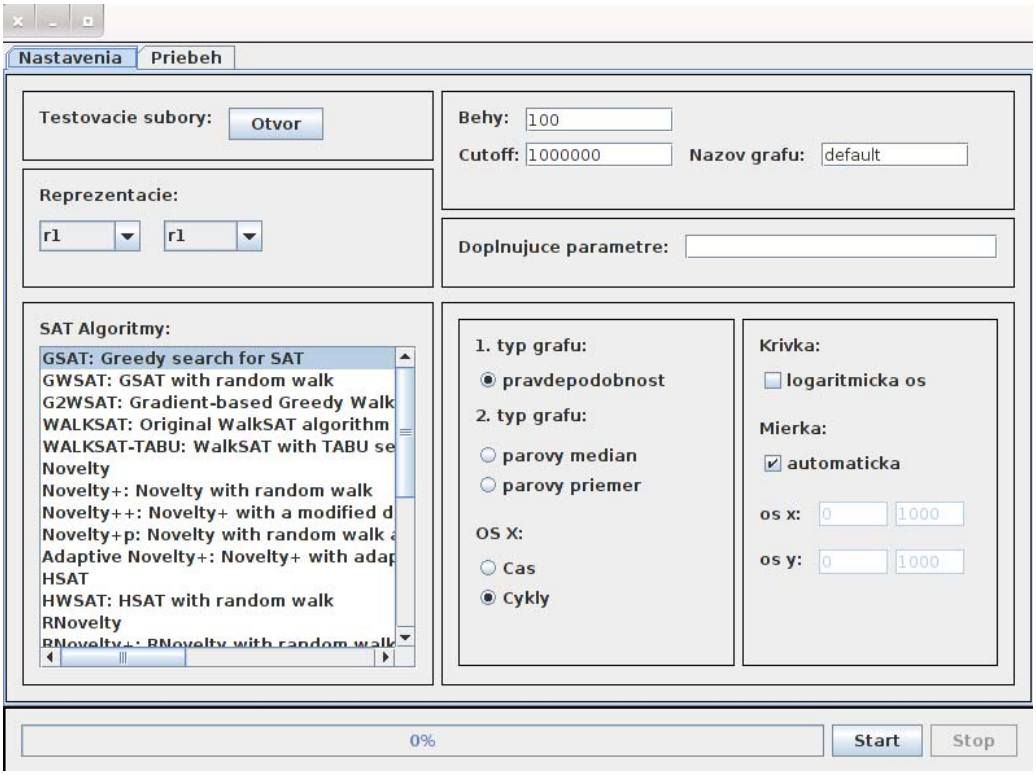
Nie všetky ohraničenia sú nutné, pretože splnenie niektorých ohraničení vyplýva zo splnenia ohraničení ostatných. Pre experimenty boli použité štyri rôzne alternatívy, reprezentujúce rôzny počet vynechaných ohraničení:

Sudoku8 – použité všetky ohraničenia (maximálna reprezentácia)
Sudoku7 – nebolo uvažované ohraničenie *b*
Sudoku6 – neboli uvažované ohraničenia *b* a *d*
Sudoku5 – neboli uvažované ohraničenia *b* a *d* a *f*

Jednotlivé reprezentácie boli postupne vytvárané z úplnej reprezentácie a následne transformované to tvaru CNF reprezentácie (pri použití binomického spôsobu kódovania ohraničenia „maximálne raz“) prostredníctvom nástroja SatSystem [5]. Pri maximálnej reprezentácii pri použití binomického spôsobu kódovania výsledný generovaný súbor obsahuje 11988 klauzúl.

B. Testovacie prostredie

Pre účely experimentovania s rôznymi reprezentáciami bol navrhnutý a implementovaný systém SatBenchmark. Jedná sa o funkčný prototyp testovacieho prostredia, umožňujúci porovnávanie viacerých reprezentácií SAT úloh na základe ich riešenia prostredníctvom vybraného algoritmu. Jeho možnosti sú zrejmé z obr. 1.

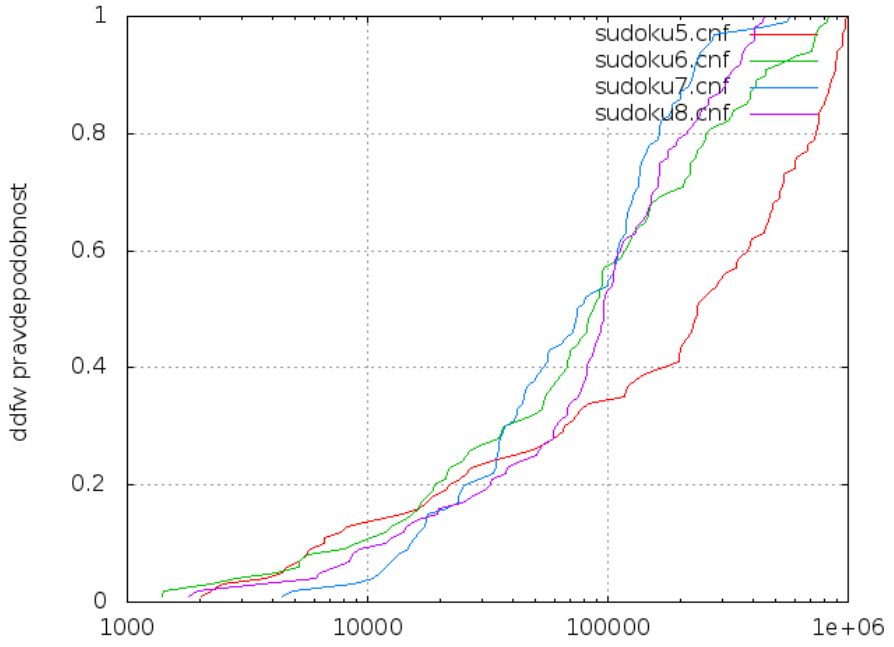


Obr.1 Testovacie prostredie SatBenchmark.

Implementovaný nástroj umožňuje párové porovnávanie rýchlosti riešenia jednotlivých reprezentácií pomocou zvoleného algoritmu ako aj porovnávanie pravdepodobností vyriešenia jednotlivých reprezentácií v limitovanom čase.

C. Dosiahnuté výsledky

Výsledky dosiahnuté jednotlivými reprezentáciami Sudoku5 až Sudoku8, prezentované v tvare závislosti pravdepodobnosti vyriešenia reprezentácie na časovom limite (reprezentovanom počtom vnútorných cyklov použitého algoritmu) sú zobrazené na obr. 2.



Obr. 2 Porovnanie testovaných reprezentácií

Interpretácia zobrazenia je jednoduchá – pri obmedzení počtu cyklov algoritmu na hodnotu 100000 bolo pri riešení reprezentácie Sudoku5 úspešných iba 35% pokusov o nájdenie riešenia

(75% pokusov vyžadovalo dlhší čas pre vyriešenie úlohy), zatiaľ čo pri ostatných reprezentáciách sa úspešnosť zvýšila na hodnoty okolo 55%. Čas s garantovaným vyriešením všetkých pokusov (dosiahnutie pravdepodobnosti rovnej 1.0) bol najkratší pre Sudoku8 a najdlhší pre Sudoku5.

Z výsledkov vyplýva, že vynechanie malého počtu ohraničení nemá dramatický vplyv na pravdepodobnosť vyriešenia úlohy v limitovanom čase. Vynechanie na nejakej hladine pravdepodobnosti môže viesť k malému zlepšeniu alebo aj k zhoršeniu. Iná situácia je pri väčšom rozsahu vynechaných ohraničení. Je očividné, že dochádza k značnému poklesu pravdepodobnosti, signalizujúcemu že použitý algoritmus riešenia má s danou reprezentáciou väčší problém.

IV. ZÁVER

Z pozorovania dosiahnutých výsledkov práce je možné vyvodiť záver, že ak jeden a ten istý problém s ohraničeniami je možné popísať viacerými spôsobmi, tak má zmysel skúmať ktoré ohraničenia do popisu problému zahrnúť a ktoré možno ignorovať (s ich implicitným splnením v prípade splnenia explicitne uvažovaných ohraničení).

Hľadaniu optimálnej reprezentácie ohraničení sa venuje napríklad [6]. V tejto práci bolo experimentálnym spôsobom ukázané, že ak na zvolený testovací problém aplikujeme iné ohraničenie tak, aby bol aj naďalej riešiteľný rovnakým spôsobom ako pri maximálnej reprezentácii, má to vplyv na rýchlosť vyriešenia daného problému. Pomocou vykonaných experimentov možno dospieť k názoru, že ak úlohu neodľahčujeme o žiadne ohraničenia, tak je vyriešená oveľa rýchlejšie ako v prípade vyhľadávania minimálnej sady ohraničení, ktorá síce umožní menej rozsiahlu reprezentáciu úlohy avšak za cenu rizika predĺženia doby potrebnej pre nájdenie riešenia.

POĎAKOVANIE

Táto práca bola vytvorená realizáciou projektu Rozvoj Centra informačných a komunikačných technológií pre znalostné systémy (kód ITMS projektu 26220120030) na základe podpory operačného program Výskum a vývoj financovaného z Európskeho fondu regionálneho rozvoja.

LITERATÚRA

[1] M. Mach, J. Paralič, *Úlohy s ohraničeniami: od teórie k programovaniu*. Elfa, Košice, 2000, 217 s.

[2] H. Hoos, T. Stuetze, *Stochastic Local Search – Foundations and Applications*. San Francisco, Morgan Kaufmann, 2005, kap. 6.

[3] A.M. Frisch, P.A. Giannaros, SAT Encodings of the At-Most-k Constraint: Some Old, Some New, Some Fast, Some Slow. In: *Int. Workshop on Modelling and Reformulating Constraint Satisfaction Problems*, 2010.

[4] I. Lynce, J. Oukaknine, Sudoku as a SAT problem. In: Proc. of the 9th *Symposium on Artificial Intelligence and Mathematics*, Fort Lauderdale, 2006.

[5] M. Lechman, *Riešenie optimalizačných úloh heuristickými metódami*. Diplomová práca, Košice, 2012, 76s.

[6] C. Sinz, Towards an Optimal CNF Encoding of Boolean Cardinality Constraints. In: *Principles and Practice of Constraint Programming*, Springer, LNCS 3709, 2005, 827-831.

Návrh a realizácia manažérskeho systému pre flexibilný montážny podnik

Pavel LINKESCH, Matej ČOPÍK, Ján JADLOVSKÝ

Katedra kybernetiky a umelej inteligencie, Fakulta elektrotechniky a informatiky,
Technická univerzita v Košiciach, Slovenská republika

pavel.linkesch@student.tuke.sk, matej.copik@tuke.sk, jan.jadlovsky@tuke.sk

***Abstrakt*— Článok sa venuje návrhu a realizácii manažérskeho informačného systému pre flexibilný montážny podnik. Využité sú softvérové prostriedky od spoločnosti Oracle, menovite Oracle Analytic Workspace a Oracle Business Intelligence Discoverer. Výsledkom je interaktívny systém, ktorého úlohou je zhromažďovať a triediť podnikové údaje z relačnej databázy podniku a ponúkať nástroje na ich analýzu pre potreby na vyšších úrovniach riadenia v rámci distribuovaného systému riadenia.**

***Kľúčové slová*— DSR, manažérsky informačný systém, Oracle, OLAP, analýza dát**

I. Úvod

Distribuovaný systém riadenia je jednou z koncepcií a prístupov k návrhu riadiaceho systému.. „Je založený na decentralizácii. To znamená, že riadenie systému je rozložené na menšie časti, ktorých riadenie prebieha lokálne a je podriadené konkrétnemu riadiacemu systému, ktorý je navrhnutý pre danú časť.“ [1] Tieto čiastkové úlohy sú navrhnuté a koordinované tak, aby dosiahli spoločný cieľ.

Pre realizáciu DSR je potrebné rozdeliť zložitý systém ma jednoduchšie časti, podsystémy, ktoré budú navzájom nezávislé. Tieto časti sa nazývajú technologické zóny. Každá zóna má určenú svoju úlohu a o jej splnenie sa stará lokálne riadenie realizované regulátorom. Úlohy sú podriadené globálnemu cieľu, ktorý má na zodpovednosti koordinátor. Ten zabezpečuje komunikáciu medzi procesmi a v prípade výpadku jedného zaistí, že systém, ako celok, bude fungovať ďalej.

II. Päťúrovňový model distribuovaného systému riadenia

Základným modelom distribuovaného systému riadenia je trojúrovňový model. Ten má riadenie hierarchicky rozdelené do troch úrovní:

- Technologická úroveň riadenia - akčné členy, snímače a ich prvotné riadenie pomocou PC alebo PLC
- Úroveň supervízneho riadenia - dispečerská úroveň - prvky SCADA/HMI
- Informačná úroveň - manažérsky softvér a internetový vizualizácia

Tento model je v praxi rozšírený firmami Allen - Bradley a Rockwell Automation do päťúrovňového modelu, kory pozostáva z týchto častí:

- Úroveň strojovej výroby
- Bázová úroveň
- Úroveň základnej bunky
- Centrálna úroveň
- Úroveň plánovania

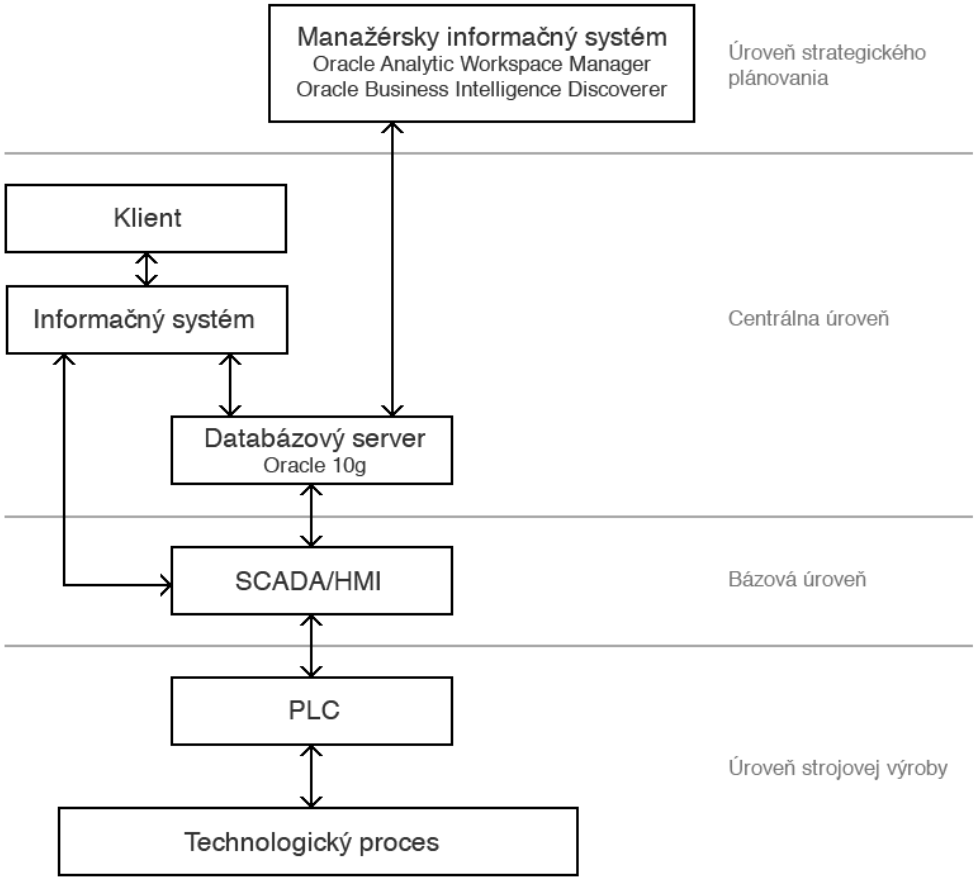
Okrem týchto základných častí model tvoria aj ďalšie doplnkové:

- Účtovníctvo
- Marketing
- Kontrola kvality
- Inžinierska činnosť

III. FLEXIBILNÝ MONTÁŽNY PODNIK

Flexibilný montážny podnik v miestnosti V147 na Vysokoškolskej ulici 4 na Katedre kybernetiky a umelej inteligencie je kompletný, samostatný výrobný podnik na báze pásovej výroby. V rámci päťúrovňového modelu distribuovaného systému riadenia sú úrovne implementované následovne [2]:

- Úroveň strojovej výroby a bázovej úrovne je tvorená PLC automatom, riadiacim počítačom a dotykovým displejom, ktoré boli programované Igorom Jadlovským v jeho diplomovej práci.
- Centrálnu úroveň tvorí informačný systém vytvorený v diplomovej práci Tomáša Krajčiho.
- Do úrovne strategického plánovania zapadá manažérsky informačný systém, ktorý je náplňou tohoto článku.
- Úroveň účtovníctva je pre potreby manažérskeho systému simulovaná údajmi v Excelovských tabuľkách zahrňujúcich informácie o spotrebe a cenách energií, nákupných nákladov na komponenty, platové ohodnotenie zamestnancov a cenník vyrobených produktov.



Obr. 1 Základná schéma flexibilného montážneho podniku [2]

Základná schéma systému je znázornená na **Obr. 1**.

IV. MANAŽÉRSKY INFORMAČNÝ SYSTÉM

Cieľom bolo navrhnúť a realizovať manažérsky informačný systém pre existujúci flexibilný montážny podnik v miestnosti V147 na Vysokoškolskej 4 na Katedre kybernetiky a umelej inteligencie.

Manažérsky systém by mal mať tieto vlastnosti a funkcie [2]:

- Zbierať dáta z nižších úrovní riadenia
- Uchovávať a spracovávať dáta
- Ponúknuť nástroje na analytickú prácu s údajmi

Všetky údaje, ku ktorým má manažérsky systém ponúknuť analytický prístup, sú zhromaždené v databáze Oracle Database 10g. Tu sa nachádzajú dáta z dvoch rôznych zdrojov

[2]:

- Výrobné údaje z výrobnej linky - To sú údaje potrebné pre beh výroby ako taký doplnené o chybové stavy.
- Údaje z informačného systému - Informačný systém do databázy ukladá informácie o užívateľoch (prístupové údaje, práva a osobné údaje), objednávkach (stav, užívateľ, zodpovedná osoba, dátumy, údaje potrebné pre výrobu a pod.) a vyrobených produktoch (kód, stav, dátum vyrobenia, pozícia v sklade, informácie o komponentoch a pod.).

Z charakteru týchto údajov vyplýva, že manažérsky informačný systém by mal poskytnúť možnosť analyzovať počty užívateľov, objednávok, produktov a chýb výrobnej linky. Tieto informácie teda budeme považovať za fakty multidimenzionálnej kocky. Keďže systémy v reálnej prevádzke v sebe zahrňujú omnoho viac údajov, ktoré by sa dali analyticky spracovávať, bolo by dobré existujúce dáta doplniť o údaje z ďalších zdrojov. Manažérsky systém, ako už názov hovorí, je určený pre potreby manažérskeho rozhodovania, pre ktoré sú nevyhnutné aj ekonomické ukazovatele. Doplníme teda databázu o údaje o cenách vyrobených produktov, nákupných cenách komponentov, o spotrebu a cenu elektriny a náklady na mzdy zamestnancov. Tieto údaje budú pre potreby práce náhodne vygenerované, takže nebudú mať reálnu výpovednú hodnotu, no doplnia pohľad na celkové možnosti systému. [2]

S ohľadom na to, že dátový model relačnej databázy bol vytvorený pre potreby informačného systému, pre zjednodušenie vytvorenia multidimenzionálnych kociek bude treba model prispôsobiť. Najvhodnejším spôsobom je nad tabuľkami v databáze vytvoriť pohľady. Takto bude zachovaná existujúca štruktúra, s ktorou informačný systém pracuje, len bude doplnená o štruktúry potrebné pre manažérsky systém. [2]

Na implementáciu vyššie popísaného návrhu riešenia a následnú analýzu dát boli zvolené dva nástroje od spoločnosti Oracle:

- Oracle Analytic Workspace Manager
- Oracle Business Intelligence Discoverer

A. Analýza dát pomocou Analytic Workspace Manager

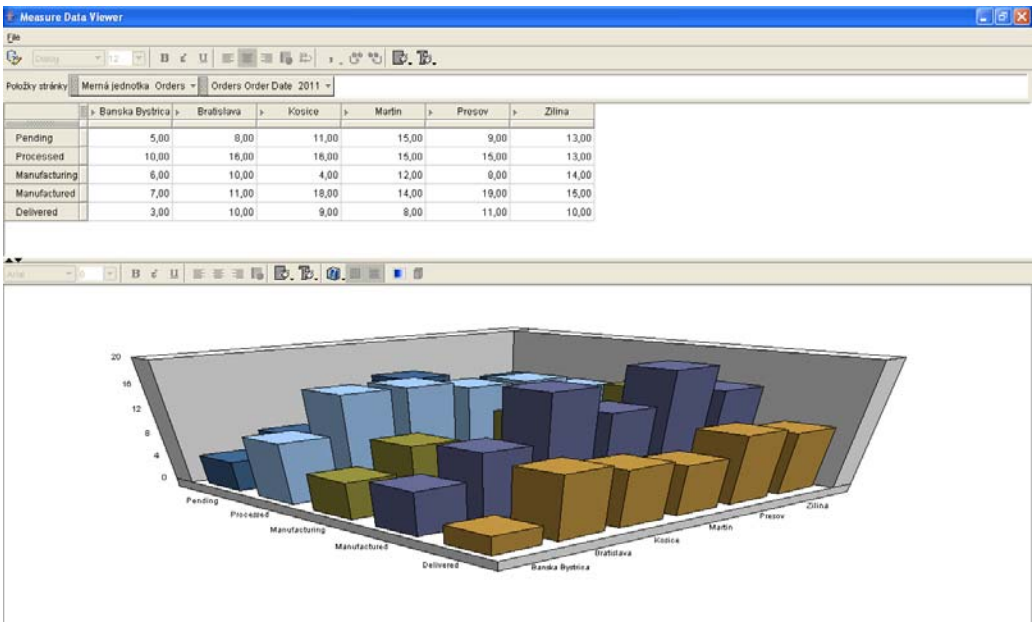
Pre prihlásenie v programe Analytic Workspace Manager použijeme nasledujúce údaje:

- Connection Information: 147.232.25.100:1521:orcl
- Username: gabriela
- Password: g4bri3l4

Bolo vytvorených 5 multidimenzionálnych kociek, ktoré pozostávajú spolu zo 14 dimenzií a 12 faktov:

- Kocka LINE_STATUS - Poskytuje analýzu chýb výrobnej linky
- Kocka ORDERS - Analýza počtu objednávok
- Kocka PRODUCTS - Analýza počtu vyrobených produktov
- Kocka USERS - Analýza počtu užívateľov
- Kocka PRICES - Analýza výšky príjmu, cien elektriny, komponentov a platov

Tieto kocky sú pripravené na analýzu. Príklad takejto analýzy vidieť na **Obr. 2**.



Obr. 2 Analýza v Analytic Workspace Manager [2]

Na **Obr. 2** vidieť počet objednávok v roku 2011 (dimenzia ORDERS_ORDER_DATE je rovná 2011). Objednávky sú rozdelené podľa ich stavov a mesta. Najviac objednávok čakajúcich na spracovanie je z Martina, najviac spracovaných z Bratislavy, najviac vo výrobe zo Žiliny a najviac vyrobených a tiež doručených v Prešove. Z grafu vidieť, že objednávok z Banskej Bystrice bolo za celý rok 2011 najmenej. [2]

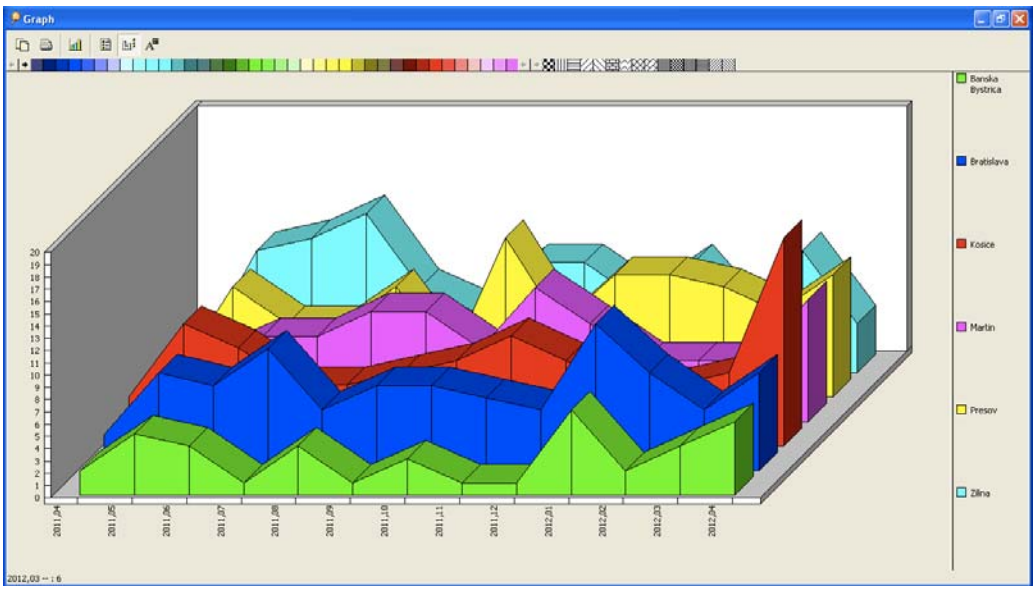
B. Analýza dát pomocou Business Intelligence Discoverer Desktop

Výhody aplikácie Business Intelligence Discoverer Desktop oproti Analytic Workspace Manager sú intuitívnejšie a komplexnejšie nástroje na prácu s údajmi a ich grafickými reprezentáciami. Na rozdiel od Analytic Workspace Manager, kde musí užívateľ stále pracovať so všetkými dimenziami, tu si môže vybrať len konkrétne stĺpce, s ktorými chce pracovať, a tak vznikajú prehľadnejšie a cielenejšie grafy.

Administrácia prebieha v programe Business Intelligence Discoverer Administrator, v ktorej bolo vytvorených 5 pracovných oblastí:

- Line Status
- Users
- Orders
- Products
- Prices

K týmto oblastiam boli pridelené prístupové práva pre užívateľa bi_user1 (heslo: bi_user1) a ten ich môže v programe Business Intelligence Discoverer Desktop využiť na analýzu. Výsledok takejto analýzy možno vidieť na **Obr. 3**, na ktorom je znázornený 3D graf, ktorý popisuje vývoj počtu objednávok v čase v jednotlivých mestách. Využitá bola pracovná oblasť ORDERS a analyzované boli stĺpce AWM_COUNT v pohľade ORDERS_VIEW a stĺpce YEAR, MONTH a DAY v ORDERS_ORDER_DATE_VIEW. Tieto dáta boli navyše doplnené o stĺpec CITY z tabuľky PERSON_DATA. Takto možno porovnať vývoj objednávok v jednotlivých mestách. Celkovo sa najslabším javí Banská Bystrica. Naopak pozoruhodný nárast vidieť v poslednom mesiaci v Košiciach. Manažérov by mohli taktiež zaujímať dôvody poklesu v Žiline v auguste 2011, keď dovtedy mal počet objednávok narastajúcu tendenciu. [2]



Obr. 3 Analýza v Business Intelligence Discoverer Desktop [2]

V. ZÁVER

Vytvorený manažérsky systém môže byť použitý pre potreby riadenia na vyšších úrovniach. Riadiacim pracovníkom umožňuje robiť rýchlejšie a lepšie informované rozhodnutia.

Použité softvérové vybavenie, programy Analytic Workspace Manager a Business Intelligence Discoverer Desktop, ponúka rozličný prístup k riešeniu toho istého problému. Prvý zmieňovaný pracuje s multidimenzionálnym dátovým modelom, takže ako prvé je potrebné vytvoriť a namapovať dimenzie, fakty a kocky. „Následná analýza dát je značne obmedzená. Analytic Workspace Manager neponúka sofistikovanejšie nástroje na prácu s kockami.“ [2] „Oproti tomu Discoverer nepracuje s multidimenzionálnymi kockami, ale s tzv. business areas, pracovnými oblasťami. Tie sa rýchlo a jednoducho naimportujú z relačnej databázy. Okrem toho Discoverer ponúka možnosť spravovať prístup viacerých užívateľov, ktorým sú podľa potreby pridelené práva pristupovať k jednotlivým pracovným oblastiam. Konečná analýza dát je oproti Analytic Workspace Manager oveľa intuitívnejšia a ponúka viac možností pre prácu s tabuľkami a grafmi.“ [2]

POĎAKOVANIE

Táto práca bola vytvorená realizáciou projektu Rozvoj Centra informačných a komunikačných technológií pre znalostné systémy (kód ITMS projektu: 26220120030) na základe podpory operačného programu Výskum a vývoj financovaného z Európskeho fondu regionálneho rozvoja.

LITERATÚRA

[1] HOŠÁK, Rastislav: Informačná úroveň riadenia modelu sústavy manipulátorov: Diplomová práca. Košice: Technická univerzita v Košiciach, Fakulta elektrotechniky a informatiky, 2009. 97 s.
[2] LINKESCH, Pavel: Návrh a realizácia manažérskeho systémy pre flexibilný montážny podnik: Diplomová práca. Košice: Technická univerzita v Košiciach, Fakulta elektrotechniky a informatiky, 2012. 71 s.

Fuzzy Cognitive Maps in Navigation of Robot Nao

Daniel LORENČÍK, Ján VAŠČÁK

Center for Intelligent Technologies, Department of Cybernetics and Artificial Intelligence, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Slovak Republic

daniel.lorencik@student.tuke .sk, jan.vascak@tuke.sk

***Abstract* — Fuzzy cognitive maps belong to the emerging approaches used for various task in artificial intelligence. They are especially useful for solving the problem of navigation of vehicle through the world, as the fuzzy systems in general are very robust, therefore suitable for real world applications. One of the disadvantages of fuzzy systems is their inability to learn. In this paper we had proposed the use of fuzzy cognitive maps for navigation of humanoid robot Nao, and also proposed adaptive mechanism for them. Data about the world are get with the use of Nao top camera. The images captured are then processed with the use of methods from Aforge framework. Depending on the situation in the arena, where is the robot, the best direction is selected with the use of membership functions for target and obstacles. Parameters of these functions can be set manually from program interface, or the optimal parameters can be found using interactive evolution. Interactive evolution was selected to obtain the best results in shortest time. Two types of creating a new generation has been used. To avoid collisions with objects outside the robot's field of view, the sonars are incorporated. Experiments were made using manually setting the parameters, using the adaptation mechanism of first and second type.**

***Keywords* — Adaptation, Fuzzy Cognitive Maps, Nao robot, Navigation,**

I. INTRODUCTION

In the bachelor thesis Fuzzy Cognitive Maps in Mobile Robotics [1] the use of fuzzy cognitive maps for solving the "Bago" problem from robot soccer was proposed. These maps were implemented not as a system of IF-THEN rules, but wit the direct use of membership functions. The space around the simulated robot was represented by angle so the objects on field were internally represented by the angle at which they were to the robot. This system of representation can be easily adapted for navigation purposes, where the obstacles and target are represented by the angle at which they are to the robot, by their relative distance to the robot and by their width. The goal of this work is also implement the adaptation system, as the fuzzy cognitive maps are unable to learn and adapt on their own. For this task, the interactive evolution was selected. With it, the parameters which influence the creation (and shape) of membership functions, are evolved. Two types of genetic algorithm are proposed, which differ in how the new generation is created. Every individual from generation is evaluated by human, because it is problematic to evaluate individual sets of parameters internally, as not only the time needed to reach the target is important. Also the distance traveled and smoothness of trajectory is important, and these properties cannot be measured without the use of overhead camera. By not relying on the use of overhead camera, application created in the process can be used in more environments.

The experiments were conducted using manual input of parameters for membership functions creation, and using adaptive systems of two types proposed.

II. FUZZY COGNITIVE MAPS

Cognitive maps were created by psychologist E. C. Tolman (in [3]) to describe how rats understand the world around them. Due to the interaction with the outside environment the image of environment is created in the brain of the living agent, which is refined by every next interaction with the environment. Cognitive maps are a very robust system which can model very complex behavior (R. Axelrod – Structure of Decision: The Cognitive Maps of Political

Elites [4]).

Fuzzy cognitive maps were created by Bart Kosko(in [2]) as fuzzy extension of cognitive maps. Fuzzy cognitive maps have clearly defined the knowledge base as every fuzzy system which is clearly separated from the structure. This allows to modify this basis without the change of structure and therefore modify the functionality. But the fuzzy cognitive maps have not the ability to learn, so this ability has to be implemented with the use of adaptive system.

Fuzzy cognitive map is an oriented graph where the nodes represent notion and the edges represent causal relations between them. The fuzziness here is implemented in weights of edges, that can be from an interval [-1; 1], and the activation values of nodes can be also from an interval [0; 1].

Fuzzy cognitive maps are a very successful method for solving complex tasks, but this also presents the problem that even the experts are not able to perceive all the causal relations. To overcome this limitation, multiple adaptation methods for FCM were proposed.

As is stated in [5], evolutionary strategies are exceptionally well suited for adapting FCM, because their main advantages are numerical optimization, robustness, adaptability and flexibility, they also use specific knowledge about problem in order to solve it. They differ from evolutionary algorithms by use of real parameters instead of binary representation. Common terms used in evolutionary strategies are population, individuals, parents and off-springs.

Individual represent the solution to given problem. Depending on the problem, the individual can be one value or have multiple parameters. Individual also represents one point n the solution space.

Population is a group of individuals in one generation.

Offspring is a new individual created either as a product of crossover between two or more parents, or as a product or mutation on one individual from population.

Parents are individuals used to create new off-springs by use of crossover. Most commonly, only the best individuals are used as parents.

Conclusion from [5] states several interesting points:

- big populations do not guarantee better results as small populations. Small populations tend to create more generations, which is more suitable for finding an optimal solution.
- process which uses the complete replacement of population is more successful than process using partial replacement
- for learning the FCM it is important to choose the right type of examples
- the success of learning is dependent on the quality and number of examples, but the dependance on the structure and complexity of FCM is not significant

III. Problem Description and Implementation

As stated before (in paragraph I.), for solving the navigation problem of robot Nao, solution proposed in [1] was used. Robot has been put to the arena seen on Figure 1, and it's task was to get to the target without collision with obstacles.

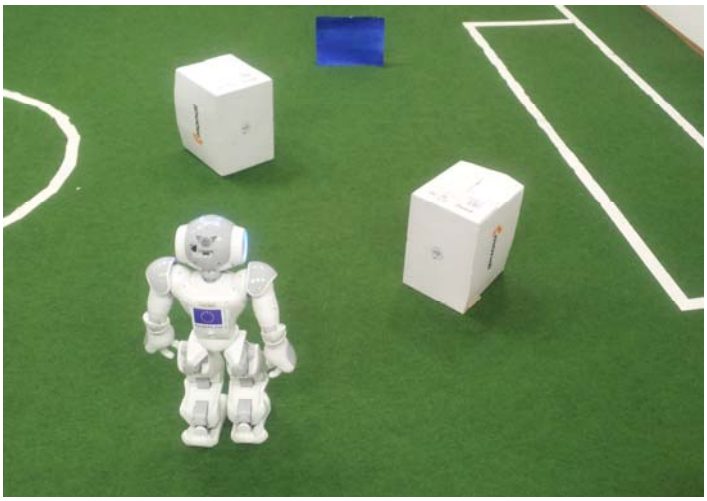


Figure 1: Arena used to adapt the FCM

The task of navigation was divided into several sub-tasks:

- communication with robot – getting images from robot camera, sending commands to the robot
- processing of the image – filtering of the image, finding object on the image
- control of the robot using information from image

A. Communication with robot

Methods from naoqi-dotnet4 library are used to communicate with robot. These methods were wrapped into classes which provides vision-related functionality, motion-related functionality and sonar-related functionality.

Getting an image of the robot is done at regular intervals in background thread, which allows for simultaneous computation. Image from robot is then stored in shared datatype, from which it can be used where needed.

B. Processing the image

For image processing, methods from Aforge framework are used. For initial filtering (to remove the arena floor from the image), Euclidean filter is used. Floor color is picked as average color of 25 pixels in the middle of image from the robot – so for getting the floor color, robot must have a clear space before it.

By using the Euclidean filter again, target is found on the image by the color. The color of the target is picked by user.

After the floor is removed, objects remaining on the image are objects in real world, which are marked and count by use of BlobCounter class from Aforge framework. Every object average color is compared to the color of biggest found target object, if there is a match, the object is marked as a target, others are marked as obstacles.

C. Representation of objects in the arena

Since the arena is perceived by robot via the visual system, only data about objects which are available are their position on the image and their size. To convert these parameters for the use of FCM, on the Figure 2 is proposed conversion to the internal coordinate system.

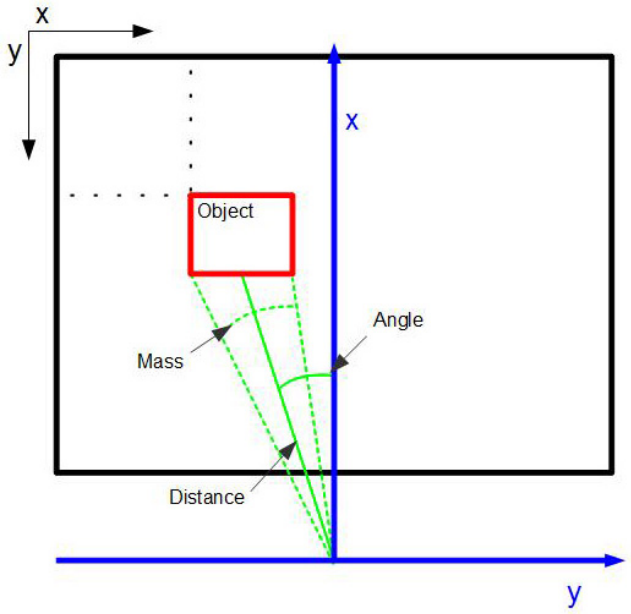


Figure 2: Conversion from image coordinate system to the internal coordinate system

After the conversion, object has following properties:

- *Angle* – angle, at which is the object regarding to robot
- *Distance* – is a relative distance between object and robot, where 1 means that object is in immediate vicinity of robot, and when the object is far, this number is lower
- *Mass* – is a width of object in degrees.

Zero point of this new coordinate system has been chosen in way, that the angle of object on image would represent angle of object to the robot in real word. For this, the situation from Figure 3 was created and angles to the objects in the corners in the field of view of robot were

measured. For object on bottom it is 45 degrees, and for the object on the top 30 degrees.

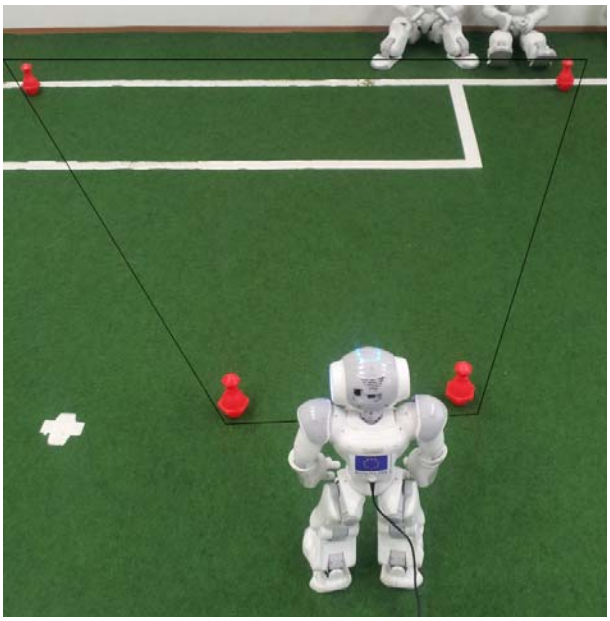


Figure 3: Field of view of robot

D. Decision process

Decision is using membership functions of obstacles and target. These functions shows dependence of validity of motion on angle of motion. The creation of function is dependent on the object parameters ans also the adaptation parameters.

The construction process consists from following steps:

- the *tangent* of function slope is calculated
- if the *tangent* is defined and not equal to zero, the length of *slope* is calculated with use of equation (1):

$$lenght\ of\ slope = \frac{distance\ of\ object * 100}{\tan(slope)} \tag{1}$$

- the *Δheight* is calculated for one step of *length of slope* (step is 1 degree):

$$\Delta height = \frac{distance\ of\ object}{lenght\ of\ slope} \tag{2}$$

- for the given range the value of function is calculated

The center of the maximum of function is located at the same angle value as object is located, the width of maximum directly corresponds to the mass of the object, and height of maximum to the distance of the object.

For target, this function is flipped – it is result of subtraction of target function from constant function with the function value of 1.

After that, all functions are summed – values of functions at same angles are summed up and the result is divided by the number of functions, so it is normalized to the interval of [0; 1].

Then the maximum of the result function shows the optimal direction for robot to reach the target. In case there are multiple maximums of the same value, higher priority is given to the maximum closer to the zero angle. Example of functions is on the Figure 4.

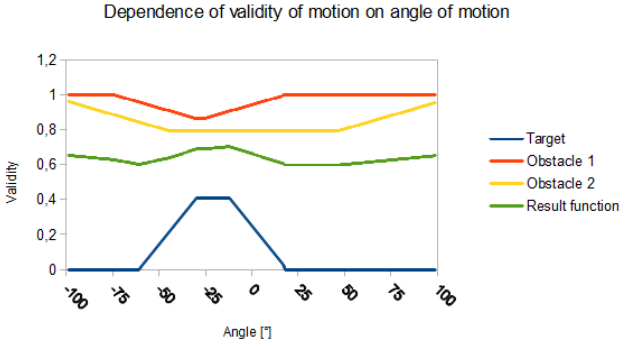


Figure 4: Example of functions used in decision process

E. Adaptation system

To adapt these functions, parameters for height, width and slope are used.

Parameter for height is from interval [0; 1] and it influences the height of the function. Height of the function is calculated as distance of the object multiplied by height parameter. This parameter is then equal to the weight that this object has on the decision.

Parameter for width is from interval [0.1; 1] and the width of the function maximum is calculated as mass of the object divided by width parameter. This will ensure that the object on the image appears wider for decision process, thus will be avoided by greater margin.

Slope parameter id from interval [0; 89] and determines the slope of the function.

These parameters can be set manually, or with the use of genetic algorithm, and they are set for obstacles and target. Also, the moving and turning speed are also set as parameters which can be optimized with the use of genetic algorithm.

The proposed solution of adaptive system uses interactive evolution to optimize the parameters. Interactive evolution was selected because the fitness function here is not clearly defined, as the fitness of one set of parameters (which we will call an individual) is determined by time it takes to reach the target, and also from trajectory of robot. For tracking the trajectory, additional camera would be required, and this process would be not possible to use in places, where the overhead camera is not present.

Individuals consist of set of parameters – height, width and slope parameters for target, height, width and slope parameters for obstacles, motion speed and turn speed. In one generation, there are 5 individuals, every one of them has to be evaluated by human operator.

Two types of genetic algorithms were used, which differ in creation of new generation.

In first type, new generation is created by using this rules:

- the best individual from old generation is copied without change to the new generation
- second new individual is a result of crossover between best and second best individual from old generation.
- third new individual is a result of crossover between second best and third best individual from old generation
- fourth new individual is a result of applying mutation on the fourth individual from old generation
- fifth new individual is a result of applying hypermutation on the fifth individual from old generation

Then the fitness of new individuals is set to zero, and they are presented to the human operator, which will evaluate them (based on performance of robot using parameters from individual).

In second type, new generation is created by using this rules:

- all possible off-springs are created by crossover (every individual produces off-spring with another individual), new individual fitness is set to the mean of fitness of it's parents
- from every off-spring, two addition individuals are created, one as a result of mutating it, and second as a result of hypermutation of off-spring
- for every parameter a condition is set. If the parameter of off-spring is in given range, the fitness of off-spring is raised by 1, otherwise is lowered by 1
- depending on fitness, 5 best off-springs are selected to the new generation
- fitness of new individuals is set to zero

Crossover used in both types creates new individual which has the parameters set as a mean of parameters of both parents.

IV. EXPERIMENTS

Table 1
Manual entry of parameters

Obstacle Height	Obstacle Width	Obstacle Slope	Target Height	Target Width	Target Slope	Motion Speed	Turn Speed
0,9000	0,8000	60,00	1,0000	1,0000	81,00	0,2000	0,5000

Experiments using manual entry of parameters and both two types of evolutionary algorithm were made in arena show in Figure 1, the best individuals from evolutionary algorithm were also tested in changed arenas. Parameters used for manual entry are in Table 1, in Table 2 are

individuals from evolutionary algorithm of first type, and in the Table 3 individuals from evolutionary algorithm of second type. In Table 4 are the conditions for evaluating off-springs in genetic algorithm of second type.

Table 2
Individuals from first type of genetic algorithm

Individual	Obstacle Height	Obstacle Width	Obstacle Slope	Target Height	Target Width	Target Slope	Motion Speed	Turn Speed	Fitness
1	0,7003	0,3906	32,00	0,1000	0,7275	52,00	0,3887	0,1927	8
2	0,4919	0,4298	83,00	0,1755	0,8937	42,00	0,8324	0,5219	7
3	0,6932	0,2593	61,00	0,1000	0,4603	9,00	0,4921	0,1000	6
4	0,0508	0,8242	31,00	0,8217	0,2465	20,00	0,8698	0,5063	6
5	0,3980	0,6535	3,00	0,2935	0,2539	58,00	0,8594	0,4534	4
1	0,7003	0,3906	32,00	0,1000	0,7275	52,00	0,3887	0,1927	8
6	0,5961	0,4102	57,50	0,1378	0,8106	47,00	0,6105	0,3573	9
7	0,2714	0,6270	57,00	0,4986	0,5701	31,00	0,8511	0,5141	8
8	0,7729	0,2176	52,01	0,1807	0,4077	10,21	0,5813	0,1000	6
9	0,4340	0,6096	10,49	0,1377	0,2160	59,05	0,7202	0,1793	10
9	0,4340	0,6096	10,49	0,1377	0,2160	59,05	0,7202	0,1793	6
10	0,5150	0,5099	34,00	0,1377	0,5133	53,02	0,6654	0,2683	7
11	0,6482	0,4004	44,75	0,1189	0,7691	49,50	0,4996	0,2750	8
12	0,1994	0,6120	55,11	0,5911	0,6000	39,21	0,7943	0,4530	6
13	0,6052	0,1620	39,36	0,2329	0,1401	2,78	0,3345	0,2934	8
13	0,6052	0,1620	39,36	0,2329	0,1401	2,78	0,3345	0,2934	6
14	0,6267	0,2812	42,05	0,1759	0,4546	26,14	0,4171	0,2842	8
15	0,5816	0,4552	39,37	0,1283	0,6412	51,26	0,5825	0,2716	9
16	0,3756	0,5953	3,29	0,1000	0,1448	55,55	0,6989	0,1681	4
17	0,1000	0,3887	43,58	0,2951	0,8215	38,23	0,5935	0,2334	6
15	**0,5816**	**0,4552**	**39,37**	**0,1283**	**0,6412**	**51,26**	**0,5825**	**0,2716**	**8**
18	**0,6042**	**0,3682**	**40,71**	**0,1521**	**0,5479**	**38,70**	**0,4998**	**0,2779**	**8**
19	**0,3634**	**0,3350**	**42,82**	**0,2355**	**0,6381**	**32,18**	**0,5053**	**0,2588**	**8**
20	**0,5126**	**0,1025**	**37,18**	**0,2580**	**0,1000**	**1,00**	**0,3702**	**0,3359**	**8**

Table 3
Individuals from second type of genetic algorithm

Individual	Obstacle Height	Obstacle Width	Obstacle Slope	Target Height	Target Width	Target Slope	Motion Speed	Turn Speed	Fitness
1	0,8536	0,2165	88,00	0,4584	0,4951	81,00	0,3475	0,3524	6
2	0,8343	0,6140	39,00	0,3487	0,3848	23,00	0,2504	0,1784	4
3	0,4395	0,5728	88,00	0,9032	0,8417	28,00	0,9041	0,3226	3
4	0,6355	0,7376	81,00	0,5243	0,6530	26,00	0,1000	0,6610	1
5	0,8325	0,8990	87,00	0,4386	0,1823	82,00	0,9133	0,8595	2
8	0,7314	0,5154	72,22	0,5581	0,1403	60,72	0,4185	0,2910	4
9	0,6466	0,3947	88,00	0,6808	0,6684	54,50	0,6258	0,3375	3
10	0,6785	0,3222	89,00	0,7372	0,7614	62,55	0,5540	0,3073	8
18	0,6369	0,5934	63,50	0,6259	0,6132	25,50	0,5772	0,2505	9
19	0,7105	0,6456	72,31	0,6516	0,5634	28,49	0,5660	0,1950	9
37	0,6890	0,6773	72,31	0,6317	0,6272	18,33	0,5127	0,3067	8
36	0,6737	0,6195	67,91	0,6388	0,5883	27,00	0,5716	0,2228	7
48	0,6945	0,4839	80,66	0,6944	0,6624	45,52	0,5600	0,2511	9
49	0,6974	0,4306	86,95	0,6212	0,6360	43,92	0,6082	0,3078	1
50	0,6254	0,5770	89,00	0,7592	0,4333	38,87	0,3417	0,2023	5
67	0,7179	0,6472	74,25	0,6160	0,7181	40,45	0,5696	0,2386	10
66	0,6917	0,5806	76,48	0,6631	0,6448	31,93	0,5364	0,2789	7
70	0,7755	0,5442	73,94	0,6601	0,5591	28,01	0,6398	0,3243	5
71	0,5412	0,7885	86,16	0,8708	0,8530	27,75	0,6238	0,1000	8
69	0,6841	0,5517	74,28	0,6666	0,6254	36,26	0,5658	0,2369	9
99	0,6295	0,7179	80,20	0,7434	0,7856	34,10	0,5967	0,1693	9
100	0,6324	0,6686	72,98	0,6804	0,7808	31,01	0,5315	0,1686	7
96	0,7010	0,5995	74,27	0,6413	0,6718	38,36	0,5677	0,2378	8
97	0,6643	0,5307	74,28	0,6329	0,7157	33,10	0,5630	0,3294	1
102	0,7048	0,6139	75,37	0,6395	0,6815	36,19	0,5530	0,2588	10

V. CONCLUSION

Results of experiments show that proposed implementation of fuzzy cognitive maps is

suitable for navigation of robot Nao. Two types of genetic algorithm were used for adaptation, the second one was able to provide optimized sets of parameters sooner than the first one, mainly due to higher number of crossovers.

Using FCM for navigation purposes has proven to be robust solution able to work even with significantly slow rate at which the images from robot can be transferred via the Wi-Fi connection.

ACKNOWLEDGMENT

This work is the result of the project implementation: Development of the Center of Information and Communication Technologies for Knowledge Systems (ITMS project code: 26220120030) supported by the Research & Development Operational Program funded by the ERDF.

REFERENCES

[1] Lorenčík, Daniel: Fuzzy kognitívne mapy v mobilnej robotike. Bakalárska práca. Košice: Technická univerzita v Košiciach, Fakulta elektrotechniky a informatiky, 2010, 44 s.
[2] KOSKO, Bart: Fuzzy cognitive maps, International Journal of Man-Machine Studies Vol 24(1), 1986, s. 65-75.
[3] TOLMAN, E. C..: Cognitive Maps in Rats and Men. The Psychological Review 55(4), 1948, s. 189-208
[4] AXELROD, R.: Structure of Decision: The Cognitive Maps of Political Elites. New Jersey, Princeton University Press 1976
[5] Koulouriotis, D. E., Diakoulakis, I. E. and Emiris, D. M. 2001. Learning fuzzy cognitive maps using evolution strategies: a novel schema for modeling and simulating high-level behavior, Proc. Of the 2001 Congress on Evolutionary Computation, Seoul, Vol. 1, pp. 364-371
[6] MLS, Karel : Fuzzy kognitivní mapy (dizertačná práca), ČVUT Praha, 2006
[7] KOTRAS, Martin: Adaptácia fuzzy kognitívnych máp pre potreby navigácie. Bakalárska práca. Košice: Technická univerzita v Košiciach, Fakulta elektrotechniky a informatiky, 2009. 45 s.
[8] PAĽA, Martin : Navigácia autonómneho vozidla pomocou fuzzy kognitívnych máp. Bakalárska práca. Košice: Technická univerzita v Košiciach, Fakulta elektrotechniky a informatiky , 2009. 41 s.
[9] RUTRICH, Marek: Adaptácia fuzzy kognitívnych máp pre potreby navigácie. Diplomová práca. Košice: Technická univerzita v Košiciach, Fakulta elektrotechniky a informatiky, 2009. 71 s.
[10] Ján Vaščák, Martin Paľa. 2012 Adaptation of Fuzzy Cognitive Maps for Navigation Purposes by Migration Algorithms. In: International Journal of Artificial Intelligence. Vol. 8, no. S12, pp. 20-37.
[11] Ivan Zelinka : Umělá inteligence v problémech globální optimalizace, BEN, Praha 2002

Robustnosť komplexných sietí a ich tolerancia k poruchám a útokom

Gabriel LUKÁČ, Marián MACH

Department of Cybernetics and Artificial Intelligence, Faculty of Electrical Engineering and Informatics Technical University of Košice, Slovak Republic

Gabriel.Lukac@tuke.sk, Marian.Mach@tuke.sk

Abstract **– Štúdium modelov robustnosti komplexných sietí a ich odolnosti voči kaskádovitým poruchám má pre rôzne odvetvia spoločenského života veľký význam. Predkladaný príspevok popisuje základné princípy robustnosti komplexných sietí. S robustnosťou priamo súvisí tolerancia sietí k poruchám a útokom. Tej je venovaná samostatná kapitola. Text ďalej uvádza numerický experiment, ktorého cieľom je overenie tolerancie vybraných modelov grafov.**

Keywords **– komplexná sieť, robustnosť, kaskádovitá porucha, náhodný graf, Barabási-Albert model.**

I. Úvod

Štúdium modelov robustnosti komplexných sietí a ich odolnosti voči kaskádovitým poruchám má pre rôzne odvetvia spoločenského života veľký význam. Na poli energetiky, dopravy, genetiky ponúka nástroje pre relatívne lacné overenie kritických situácií, ktoré môžu negatívne ovplyvniť stabilitu zložitých systémov. V nasledovnom texte budú načrtnuté teoretické základy robustnosti komplexných sietí a princípy niektorých modelov relevantnývch k adresovaniu zmienených problémov.

II. Robustnosť

Podľa [1] **robustnosť** siete spočíva v jej schopnosti predchádzať situáciám, kedy prestáva plniť funkciu, na ktorú bola navrhnutá. Tento stav môže nastať vtedy, ak sa niektorá jej časť (závisí to od jej významnosti v rámci topológie a aktuálne spustených procesov) poškodí. Adresovanie tohto problému má praktické dôsledky pretože priamo ovplyvňuje odolnosť systému voči destabilizácii.

Pri popisovaní robustnosti sietí je potrebné rozlíšiť dva prístupy: prvý je **statická** robustnosť, druhým **dynamická**.

A. Statická robustnosť

Statická robustnosť referuje na činnosť, kedy sa zo siete odstraňujú uzly (a súčasne aj s nimi incidujúce hrany) bez akejkoľvek potreby redistribuovať kvantitu, ktorá sa cez túto sieť transportuje. To sa môže týkať napríklad prípadu sociálnych sietí, kedy sa zo siete „vymažú“ konkrétni participanti a takto sa prerušia vytvorené väzby. Na druhej strane dynamická robustnosť odkazuje na potrebu vziať do úvahy dynamiku predistribuovania kvantity „tečúcej“ porušenou sieťou. Typickým príkladom môže byť sieť Internet, kedy pri výpadku smerovača musia byť pakety presmerované alternatívnymi vetvami.

Načrtnuté prípady robustnosti vyzerajú navonok podobne, avšak zatiaľ čo statická robustnosť môže byť riešená analytickými metódami (napr. perkolačnou teóriou [2] majúcou korene v štatistickej fyzike), ten druhý prípad je oveľa ťažší a takmer vo všetkých prípadoch je potrebné spoliehať sa na numerické simulácie.

B. Dynamická robustnosť

Ak sa chceme priblížiť k úplnému popisu robustnosti systému so sieťovou štruktúrou, nemôžeme ostať iba pri popise statických vlastností. Pre úplnosť charakteristiky je potrebné pridať do popisu dynamické aspekty a vytvoriť tak jeden model zachytávajci tak statické ako aj dynamické vlastnosti modelovaného problému. Pri modelovaní reálneho systému však takmer vždy narážame na vysokú zložitosť keďže rôzne komponenty siete majú rôzne prenosové kapacity (prenosová priepustnosť) a zaťaženosť jednotlivých uzlov (alebo spojení) je premennou s vysokou variabilitou čo do zmeny časovej ale aj priestorovej [1]. Typickým príkladom sú tzv. *kaskádovité poruchy* (angl. cascading failures) spôsobujúce lavínové preťaženia častí siete.

III. Tolerancia k poruchám a útokom

Dôvody na študovanie robustnosti sietí k neočakávaným výkyvom konektivity v podobe náhodných porúch alebo cielených útokov je možné charakterizovať nasledovnými dvoma bodmi [3]:

1) Návrh nových sietí ako systémov integrovaných do ich prostredia, berúc do úvahy, či sa pri navrhovanom dizajne a prostredí môžu vyskytnúť len náhodné poruchy, alebo aj cielené útoky.
2) Ochrana existujúcich sietí identifikáciou najkritickejších uzlov a prijatie protiopatrení na redukciu ohrozenia.

Siete sa vo všeobecnosti vyznačujú vysokou mierou robustnosti, ktorá sa nezníži ani pri odstránení konkrétneho množstva hrán alebo uzlov. Prvé výsledky študovania robustnosti boli získané analýzou a experimentovaním s náhodnými grafmi [4][5]. Tie uvažovali ľubovoľný graf H_N s N vrcholmi spolu s predpokladom, že z grafu bol zmazaný zlomok vrcholov veľkosti p zo všetkých možných. Aká je pravdepodobnosť, že výsledný podgraf je súvislý a ako to závisí od pravdepodobnosti zmazania zlomku p? Pre širokú triedu grafov H_N existuje prahová pravdepodobnosť $p_c(N)$ taká, že ak $p < p_c(N)$, potom výsledný podgraf je súvislý, ale ak $p > p_c(N)$ výsledný podgraf je nesúvislý.

Podobná situácia nastáva pri zmazaní uzla, čím nevyhnutne dôjde aj k zmazaniu nejakého počtu hrán a preto sa dá povedať, že zmazanie uzlov implikuje vyššie škody ako odstraňovanie hrán. V nasledovnom texte je uvedený krátky prehľad numerických výsledkov o hľadaní prahovej hodnoty pre pravdepodobnosť zmazania vrchola a jej dopad na robustnosť siete. Výskumy [6] dokazjú, že existuje silná korelácia medzi topológiou siete a jej robustnosťou. Ak by sme chceli byť konrétni, podľa zistení horeuvedených autorov, scale-free siete sú oveľa odolnejšie voči náhodným poruchám ako náhodné grafy. Na druhej strane sú zase zraniteľné voči cieľeným útokom na uzly s veľkým počtom hrán.

Autori v [6] skúmali siete rovnakej veľkosti (počet uzlov a hrán bol rovnaký), líšili sa iba v distribúcii stupňov vrcholov. Boli uvažované dva spôsoby odstránenia uzlov, náhodné (ktoré simulovalo náhodné pertrubácie v sieti majúce v reálnych situáciách za následok výpadok zasiahnutých uzlov) a druhé, spočívajúce v postupnom odstraňovaní uzlov s vysokým stupňom, emulujúce zámerný útok.

Numerický experiment bol nasledujúci: Začnime so súvislou sieťou a v každom kroku odstráňme jeden uzol. Zmazanie uzla má za následok súčasné odstránenie všetkých s ním incidentných hrán a zrušenie niektorých ciest medzi dvojicami uzlov. Jedným zo spôsobov ako sledovať porušenie siete je pozorovanie relatívnej veľkosti najväčšieho zhluku S, ktorý zostáva súvislý a priemernej hodnoty dĺžky cesty l daného zhluku ako funkcie podielu odstránených uzlov. Očakáva sa, že veľkosť najväčšieho súvislého zhluku sa bude s ubúdajúcimi uzlami zmenšovať, naopak priemerná dĺžka cesty medzi dvoma vrcholmi bude narastať [7].

a) Náhodný graf, náhodné odstránenie uzla: Pri pokuse s náhodným grafom s parametrami $N = 10000$ a $\langle k \rangle = 4$, veľkosť najväčšieho zhluku S sa znižuje z hodnoty $S = 1$ ak sa f zvyšuje, kde S je relatívna veľkosť najväčšieho zhluku a f je zlomok (percento) odstránených uzlov. Pri dosiahnutí kritickej hodnoty f_c, S sa znížlo na 0 indikujúc, že sieť sa rozbila na malé izolované zhluky. Čo sa týka hodnoty priemernej vzdialenosti l medzi uzlami, jej rast začína z hodnoty charakteristickej pre neperturbovanú náhodnú sieť a rastie spoločne s f tak, ako sú v sieti zmazávané jednotlivé hrany. Jej hodnota kulminuje pri kritickej hodnote f_c. Po prekročení tejto hodnoty dochádza v sieti k „rozbitiu“ veľkých zhlukov na malé (teda aj najväčší zhluk sa prudko zmenšuje) a hodnota l sa znižuje rapídne.

b) Scale-Free sieť, náhodné odstránenie uzla: Zatiaľčo pri náhodnom grafe dochádza k inverznej perkolácii pri náhodnom odstránení uzla, pri modeli Barabási-Albert (BA) je situácia úplne iná. Simulácie ukazujú, že veľkosť najväčšieho zhluku sa znižuje a dosahuje hodnotu 0 pri vyššom f ako to bolo v prípade náhodných grafov. Súčasne l rastie oveľa pomalšie ako v prípade náhodných grafov a tvar grafu v bode jeho kulminácie nie je až taký jednoznačný ako pri Erdös-Rényi modeli. Správanie sa systému stále naznačuje perkolačný prechod, ale analytické výpočty ukazujú, že toto je efekt iba pre siete konečnej veľkosti. Inými slovami sa dá povedať, že scale-free sieť vykazuje výnimočnú robustnosť voči náhodnému odstráneniu uzlov.

c) Preferenčné odstránenie uzla: V tomto prípade cielené odstraňovanie uzlov vysokého stupňa vedie k rýchlejšiemu porušeniu siete ako keby sme podobný pokus vykonali na náhodných

grafoch. Všeobecný scenár „pokazenia sa“ siete sleduje inverzný perkolačný prechod[1], ale kritický pomer je oveľa nižší ako pri náhodnom odstraňovaní. To je celkom pochopiteľné, keďže v každom z krokov sa zo systému odstráni spolu s vrcholom vysokého stupňa aj relatívne vysoký počet hrán naraz. Dopad tohto procesu má na siete s rôznou topológiou rôzny dopad. Scale-free sieť sa rozpadne na nesúvislé komponenty rýchlejšie (keďže jej štruktúra do veľkej miery závisí od uzlov s vysokým stupňom), než je to v prípade náhodného grafu.

Analytické riešenie horeuvedeného problému robustnosti sietí pomocou perkolačnej teórie je možné nájsť v prácach [4] alebo [7].

Iný prístup k riešeniu problému objavenia sa náhodných chýb a útokov zaujali autori [3], ktorí v simulácii odstraňovali uzly cielene na základe troch rôznych kritérií:

1) *Stupňa vrchola*, t.j. na základe počtu incidujúcich hrán s daným vrcholom,
2) *betweenness miery* alebo *zaťaženia uzla*, napr. počet najkratších ciet medzi dvoma uzlami (spomedzi všetkých najkratších ciest v sieti) prechádzajúcich daným uzlom, záťaž sa v tomto prípade vyhodnocuje pred odstránením uzla,
3) *prepočítaná betweenness miera* alebo *prepočítaná záťaž*, t.j. rovnaká miera ako v predchádzajúcom prípade, s tým rozdielom, že najkratšie cesty sú prepočítané zakaždým, keď je uzol odstránený.

Pre evaluáciu svojho prístupu autori používajú mieru tzv. *globálnu efektívnosť siete* [9]. Vychádza z reprezentácie siete pomocou matice susednosti s tým, že prvok matice $a_{ij} = 1$, ak v sieti existuje hrana z uzla i do uzla j, inak je jeho hodnota nastavená na ∞. Hodnota a_{ij} (nemusí to byť len číslo 1, ale aj reálne číslo pri ohodnotenej sieti) reprezentuje čas, ktorý trvá prejdeniu informácie z uzla i do uzla j. Ak medzi týmito uzlami neexistuje priama cesta (t.j. nie sú spojené hranou), potom je informácia medzi nimi komunikovaná po najkratšej možnej ceste. Hodnotu najkratšej možnej cesty označme ako t_{ij}. Globálna efektívnosť siete je potom definovaná ako priemerná hodnota efektívností $\varepsilon_{ij} = 1/t_{ij}$ cez všetky dvojice uzlov [9]:

$$E(G) = \frac{\sum_{i \neq j \in G} \varepsilon_{ij}}{N(N-1)} = \frac{1}{N(N-1)} \sum_{i \neq j \in G} \frac{1}{t_{ij}}. \quad (1)$$

Táto miera je potom v článku použitá na vyhodnotenie tolerancie voči poruchám a útokom, konkrétne pri evaluácii experimentov vykonaných pomocou ER a BA modelov s celkovo $N = 2000$ uzlami a $K = 10000$ hranami, pričom experimentálne ukázali, že postačuje cielene odobrať 15% uzlov zo scale-free siete aby jej efektívnosť klesla na polovicu počiatočnej hodnoty. Ak zámerne odoberieme 35% uzlov, dochádza k úplnému zničeniu siete. Naopak autori ukázali, že pri náhodnom odoberaní uzlov (pri odobraní 80%) si scale-free sieť vie zachovať relatívne slušnú mieru efektívnosti, $E = 0.15$. Pri odstraňovaní uzlov z ER modelu, štruktúra založená na náhodnom grafe vykazuje pri náhodnom odoberaní uzla podobné vlastnosti ako BA model. Pri zámernom útoku na konkrétne uzly je pre podobné zníženie efektívnosti ako u BA modelu potrebné pre ER sieť odstrániť približne o 20% uzlov viac.

IV. Záver

Predkladaný príspevok zhrnul súčasný stav problematiky robustnosti komplexných sietí. V rámci textu boli zhrnuté základné princípy robustnosti ako aj načrtnutá tolerancia vybraných modelov grafov k poruchám a útokom. Cieľom textu je poskytnúť prípadnému záujemcovi prehľad súčasného stavu a zoznam relevantných zdrojov pre hlbšie štúdium.

Poďakovanie

This work is the result of the project implementation: Development of the Center if Information and Communication Technologies for Knowledge Systems (ITMS project code: 26220120030) supported by the Research & Development Operational Program funded by the ERDF (100%).

[1]Perkolačná teória majúca korene v štatistickej fyzike ponúka aparát pre popísanie fázových prechodov. Autori [8] definujú perkolačný proces ako šírenie tekutiny médiom, pričom dané médium definuje náhodný mechanizmus, ktorý šírenie ovplyvňuje. Ako médium sa používa štvorcová mriežka bodov v rovine (fyzici ju nazývajú štvorcový zväz, angl. square lattice). Cieľom perkolačného procesu je „prekvapkanie“ tekutiny z jedného konca mriežky na druhý, pričom tekutina sa šíri iba uzlami, ktoré sú špeciálne zafarbené. Každý uzol mriežky je zafarbený s pravdepodobnosťou p, pričom tá je pre každý uzol nezávislá od toho, či už sused zafarbeného uzla bol alebo nebol zafarbený. To znamená, že ak má sieť N uzlov, pričom N je veľmi veľké číslo, potom pN uzlov bolo zafarbených a $(1-p)N$ nie. Cieľom perkolačnej teórie je nájsť takú hodnotu kritickej pravdepodobnosti p_c, pri ktorej v nekonečnej mriežke vznikne nekonečne veľký zhluk zafarbených uzlov.

LITERATÚRA

[1] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D. Hwang, "Complex networks: Structure and dynamics," *Physics Reports*, vol. 424, no. 4-5, pp. 175–308, 2006.
[2] D. Stauffer and A. Aharony, *Introduction To Percolation Theory*. CRC, July 1994.
[3] P. Crucitti, V. Latora, M. Marchiori, and A. Rapisarda, "Error and attack tolerance of complex networks," *Physica A: Statistical Mechanics and its Applications*, vol. 340, no. 1-3, pp. 388 – 394, 2004, news and Expectations in Thermostatistics. [Online]. Available: http://www.sciencedirect.com/science/article/B6TVG-4C8DCSG-8/2/d89540521d4bc93ffe29486484881e24
[4] D. S. Callaway, M. E. J. Newman, S. H. Strogatz, and D. J. Watts, "Network robustness and fragility: Percolation on random graphs," *Phys. Rev. Lett.*, vol. 85, no. 25, pp. 5468–5471, Dec 2000.
[5] B. Bollobás and O. Riordan, "Robustness and vulnerability of scale-free random graphs," *Internet Mathematics*, vol. 1, no. 1, 2003.
[6] R. Albert, H. Jeong, and A.-L. Barabasi, "Error and attack tolerance of complex networks," *Nature*, vol. 406, no. 6794, pp. 378–382, July 2000.
[7] R. Albert and A.-L. Barabasi, "Statistical mechanics of complex networks," *Reviews of Modern Physics*, vol. 74, no. 1, pp. 47–97, 6 2001.
[8] S. R. Broadbent and J. M. Hammersley, "Percolation processes. I. Crystals and Mazes," in *Proceedings of the Cambridge Philosophical Society*, ser. Proceedings of the Cambridge Philosophical Society, vol. 53, Jul. 1957, pp. 629–641.
[9] V. Latora and M. Marchiori, "Efficient behavior of small-world networks," *Phys. Rev. Lett.*, vol. 87, no. 19, p. 198701, Oct 2001.

Aplikácia dolovania dát vo výrobnom priemysle

Alexandra LUKÁČOVÁ

Katedra kybernetiky a umelej inteligencie, Fakulta elektrotechniky a informatiky,
Technická univerzita v Košiciach, Slovenská republika

alexandra.lukacova@tuke.sk

***Abstrakt* — V časoch, keď ekonomiku charakterizuje rýchly a neustále sa meniaci trh, je pre podnik zložité rozpoznať ich dynamiku a zareagovať v čo najrýchlejšom čase. Tento tlak je cítiť aj vo výrobnom sektore, kde je taktiež potrebné získať konkurenčnú výhodu a udržať si tak starého alebo prilákať nového zákazníka. Preto sa aj vzhľadom k pokrokom v systémoch zameraných na zber dát stáva čoraz viac samozrejmosťou aplikovať dolovanie dát a získať tak doposiaľ neznáme znalosti, ktoré pomôžu vylepšiť výrobný proces. Táto práca prezentuje riešenie problému chemickej spoločnosti vyrábajúcej polyamidové vlákna, ktorá má záujem spoznať, ktoré vlastnosti majú vplyv na výslednú kvalitu vlákna.**

***Kľúčová slová*—dolovanie dát, výroba, kvalita, klasifikácia**

I. ÚVOD

Nezáleží na tom, o aký druh alebo veľkosť priemyslu sa jedná, každý podnik sa neustále pokúša získať konkurenčnú výhodu nad svojimi súpermi. Hlavne v týchto časoch, keď ekonomiku charakterizuje rýchly a neustále meniaci sa trh, je pre podnik rozpoznať meniacu sa dynamiku a zareagovať v čo najrýchlejšom čase, značne zložité. Preto sa vzhľadom k pokrokom v systémoch zameraných na zber dát stáva čoraz viac samozrejmosťou aplikovať techniky objavovania znalostí (KDD – Knowledge Discovery in Databases). KDD predstavuje iteratívny poloautomatický proces objavovania a verifikácie vzorov v obrovskom množstve údajov, a to práve využitím technológií vyhľadávania vzorov a štatistických a matematických techník [4].

Dolovanie dát, jeden krok procesu KDD, bol pôvodne určený len marketingovým oddeleniam ako riešenie pre dosiahnutie zvýšenia odbytu, no teraz sa úspešne aplikuje aj v iných úlohách, ako je napríklad odhaľovanie podvodov alebo skórovanie solventnosti klienta. Vyššie uvedené oblasti však nie sú všetky. Tlak konkurencie je možné cítiť aj vo výrobnom priemysle a tak je nutnosťou neustále zlepšovanie výkonnosti výroby, ktorú možno využiť k obchodnej i výrobnej úspešnosti. Jednou z možností, ako to dosiahnuť, je zamerať sa na dosiahnutie čo najlepšej kvality výsledného produktu a odlíšiť sa tým od konkurentov.

Existuje viacero analýz, ktoré sa venovali tejto téme z viacerých pohľadov. Koksal v štúdii [3] preto uvažuje o nasledovnom rozdelení úloh: popis kvality produktu alebo procesu kvality, predikcia kvality, klasifikácia kvality a optimalizácia parametrov. V našej práci sa budeme zaoberať práve klasifikáciou. Využijeme pritom rozhodovacie stromy, ktoré sú obľúbené pre ich jednoduchú interpretáciu a identifikujeme tiež atribúty najviac vplývajúce na výstup – výslednú kvalitu.

II. ANALÝZA SÚČASNÉHO STAVU

Využívanie dolovania dát vo výrobnej oblasti začalo začiatkom 90-tych rokov. Irani [2] vtedy vo svojej práci generalizoval algoritmus, ktorý dokázal predikovať výstup budúcich experimentov za rôznych, všeobecných podmienok. Jeho systém bol úspešne použitý v rôznych polovodičových priemyselných aplikáciách.

V [5] sa autori štúdie zaoberali identifikovaním najvplyvnejších premenných, ktoré spôsobili chybovosť výrobku. Liatinová turecká spoločnosť im poskytla údaje za obdobie piatich mesiacov pre výrobok s vysokým percentom chybovosti. Spoločnosť zachytávala jeho hodnoty v hodinových, denných ale aj týždenných intervaloch, no bez ich akejkoľvek následnej analýzy. Do úvahy sa vzali iba dve najdôležitejšie a často pozorované typy defektov. Z tohto

dôvodu cieľový atribút nakoniec obsahoval tri nominálne hodnoty (0 – ani jedna z chýb nebola pozorovaná, 1 – chyba č. 1, 2 – chyba č. 2). Použitím algoritmu C5.0 sa vygenerovalo desať pravidiel, ktoré spoločnosti poskytli pridanú hodnotu.

V [13] autori predstavili aplikáciu dolovania dát z údajov o výrobe tkanín poskytnutých tureckou firmou špecializujúcou sa na výrobu kobercov. Zložitosť ich vypracovanej analýzy chybovosti výrobkov spočívala v príliš mnohých kategóriách predikovaného atribútu. Na zredukovanie počtu kategorických premenných použili analýzu dôležitosti atribútov na základe informačného zisku. Nakoniec pomocou algoritmu C4.5 - rozhodovacieho stromu navrhli model, z ktorého bolo vygenerovaných 101 pravidiel. Jeho presnosť predstavovala 72,811%.

III. EXPERIMENTY

Nižšie popísanú aplikáciu sme riešili v zmysle metodológie CRISP-DM [8] (Cross Industry Standard Process for Data Mining), pozostávajúcej zo šiestich fáz – pochopenie problému, pochopenie dát, predspracovanie dát, modelovanie, ohodnotenie a nasadenie. Tejto metodológii je taktiež prispôsobený nami použitý nástroj na dolovanie dát SPSS Clementine 10.1, ktorý bez nutnosti programovania pomáha zostavovať predpovedné modely.

A. Pochopenie problému

Chemický podnik zaoberajúci sa výrobou vláken považuje svoju vstupnú surovinu – granulát za základný kameň bezporuchového procesu výroby kvalitného produktu. Aj keď celý proces výroby vlákna charakterizujú nasledovné čiastkové procesy

- proces sušenia – výstupom je vysušený granulát,
- proces zvlákňovania – výstup tvoria vlákna roztaveného granulátu,
- proces navíjania – výstup predstavuje cievka s navitými vláknami,
- proces testovania – výstupom je certifikát na zákale otestovanej vzorky,
- proces balenia – výstupom sú zabalené krabice obsahujúce cievky,

v práci sa sústredíme na identifikáciu významných premenných vplývajúcich na výslednú kvalitu výrobku, ktoré sú merané ešte pred vstupom suroviny do výrobného procesu. Keďže cieľom podniku je redukovať výrobu vláken s nízkou kvalitou, a zvýšiť tak efektívnosť výroby, má záujem spoznať, ktoré vlastnosti, a či vôbec niektoré, majú vplyv na výslednú kvalitu vlákna. Takýmto spôsobom chce následne dodávateľom stanoviť kritéria vlastnosti dodávaného granulátu.

B. Pochopenie a predspracovanie dát

Pretože je užitočné pred predspracovaním dát získať na dáta celkový pohľad, predchádza jej fáza pochopenia dát. Tá predstavuje podrobné oboznámenie sa s dátami, zistenie ich kvality a charakteru [9]. Poskytnutá množina dát pozostáva z 240 pozorovaní popísaných 13 atribútmi (X1-X13) zozbieranými počas štvormesačnej periódy. Tieto atribúty môžeme rozdeliť do dvoch kategórií. Prvých osem opisuje vlastnosti granulátu pred vykonaním post-kondenzácie, ďalších päť je nameraných po nej. Posledný atribút je cieľový atribút a charakterizuje ho percento prvej kvality vyrobeného vlákna. Všetky spomenuté atribúty sú numerické.

Fáze predspracovania sa prikladá najväčší význam a z celého procesu objavovania znalostí zaberá aj najviac času. Je však naozaj dôležité venovať jej najviac pozornosti pretože nízka kvalita dát veľmi často prináša aj nízku kvalitu výsledných modelov.

Na začiatku sme odstránili duplicitné prípady (rovnaké ID), a to z dôvodu, že sme nevedeli určiť, ktorý z dvojice bol nameraný správne (1 odstránená dvojica). Následne sme taktiež odstránili prípady s chýbajúcimi hodnotami (21 v cieľovom atribúte, 3 v atribúte X3). Nakoniec sme pristúpili ku kontrole jednotlivých atribútov a vzťahov medzi nimi. Po tejto analýze sme sa rozhodli neuvažovať s piatimi atribútmi. Dva z nich boli merané duplicitne kvôli kontrole vlastností a mali na výslednú kvalitu rovnaký vplyv (X4-X5). Tri prejavili medzi sebou vysokú koreláciu (0,999) a aj počas konzultácií nám expert potvrdil, že tieto atribúty sú len podmnožinou iného atribútu (X6, X12, X13).

Cieľový atribút je síce numerického charakteru, úlohu sme sa ale rozhodli riešiť klasifikáciou dvoch tried – nízka (<=92%) a vysoká kvalita vlákna (>92%).

Diskretizačné algoritmy hrajú pri dolovaní dát dôležitú úlohu, a to i v našom prípade, keď algoritmy rozhodovacích stromov dokážu pracovať s numerickými hodnotami. Nielen že vykonávajú stručnú sumarizáciu atribútov, čím odborníkom uľahčujú ľahšie interpretovať dáta, ale taktiež umožňujú učiť sa viac presnejšie a rýchlejšie [11]. My sme sa rozhodli pre využitie techniky chi zlučovania. Chi zlučovanie [11] využíva χ^2 štatistiku na určenie, či relatívne triedne početnosti susediacich intervalov sú výrazne odlišné, alebo sú dostatočne podobné na

ich zlúčenie do jedného intervalu. χ^2 je štatistická miera používaná na testovanie hypotéz, či dve diskrétne hodnoty sú štatisticky nezávislé. V tomto prípade to znamená, že sa testuje hypotéza, či triedny atribút je nezávislý od dvoch susediacich intervalov. Keď je záverom, že trieda je od nich nezávislá, intervaly sa zlúčia. Ak ale χ^2 test potvrdí, že nie sú nezávislé, indikuje to, že rozdiel v relatívnych triednych početnostiach je štatisticky významný a preto by mali zostať intervaly oddelené. Týmto spôsobom sme pomocou voľne dostupného štatistického softvéru R [6], a jeho rozšírením o balíček „discretization" [7], diskretizovali všetky vstupné numerické atribúty (hladina významnosti bola stanovená na 0,05).

Posledný problém nastal pri pohľade na nevyváženú distribúciu triedneho atribútu. Už v [10] sme poukázali na to, že model nedokáže klasifikovať vysoko kvalitné vlákna. Množine dát s dvomi triedami sa hovorí nevyvážená, keď jedna z tried (minoritná časť) je v porovnaní s druhou triedou (majoritná časť) reprezentovaná veľmi malým množstvom príkladov. Naučené modely, neberúc do úvahy distribúciu tried, majú snahu dosiahnuť čo najvyššiu úspešnosť. Minoritná trieda je pritom zvyčajne z pohľadu učenia sa najdôležitejšou časťou. Preto je potrebné danú distribúciu tried vybalancovať. Toto je možné dvomi spôsobmi - zvýšením počtu vzoriek minoritnej (pozitívnej) triedy, alebo znížením vzoriek majoritnej (negatívnej) triedy (jeho nevýhodou je ale potenciálna strata hodnotných prípadov). Obidve techniky taktiež vedú k zníženiu presnosti negatívnej triedy [1]. Najjednoduchšou metódou v našom prípade, aj vzhľadom na nízky počet pozorovaní, je zvýšenie počtu prípadov minoritnej triedy náhodnou replikáciou pozitívnych príkladov (faktor=3,7), a to bez pridania novej informácie do dátovej množiny. Experimentálne výsledky v práci autora Garciu [1] taktiež ukázali, že vyrovnanie distribúcie pridaním príkladov vedie k vyššej úspešnosti modelu, než použitie techniky odoberania príkladov pre dátové množiny, kde je nerovnováha tried veľká.

C. Modelovanie a vyhodnotenie

Na modelovanie sme pre jednoduchú interpretovateľnosť a ľahšiu pochopiteľnosť zvolili algoritmy rozhodovacích stromov (C5.0, QUEST, CHAID, C&R Tree). Už po prvotných experimentoch sme najlepšie výsledky dosahovali pri učení algoritmom C5.0, preto sme sa rozhodli ďalej pracovať len s ním. Ten je rozšírením známeho algoritmu C4.5, využívajúceho ako deliace kritérium pomerový informačný zisk. Pri nastaveniach sme nastavili využitie boostingu a 10 násobnej krížovej validácie.

Results for output field class

Comparing $C-class with class

'Partition'	2_Testing	
Correct	51	80,95%
Wrong	12	19,05%
Total	63	

Coincidence Matrix for $C-class (rows show actuals)

'Partition' = 2_Testing	1	2
1	10	5
2	7	41

Obr. 1 Výsledná matica úspešnosti pre testovaciu množinu

Pre vyhodnotenie úspešnosti sme predspracované dáta rozdelili na trénovaciu (70%) a testovaciu (30%) množinu. Následne sme po vytvorení modelu sledovali nasledujúce tri štatistické miery: presnosť, citlivosť a špecifickosť. Tie sú definované pomocou hodnôt, ktoré určujú, či je prípad správne alebo nesprávne klasifikovaný. Keď je prípad pozitívny a je klasifikovaný pozitívne, patrí do skupiny správne klasifikovaných (TP - true positive), keď je klasifikovaný nesprávne, patrí do skupiny nesprávne klasifikovaných (FN – false negative). Keď je prípad negatívny a je aj klasifikovaný negatívne, patrí do skupiny správne klasifikovaných (TN - true negative), keď je klasifikovaný nesprávne, patrí do skupiny nesprávne klasifikovaných (FP – false positive). Presnosť nášho modelu, vypočítaná ako

$$\frac{TP + TN}{N}, \tag{1}$$

dosiahla úspešnosť 80,95%. Citlivosť, vyjadrujúca pomer správne klasifikovaných pozitívnych prípadov oproti všetkým pozitívnym prípadom

$$\frac{TP}{TP+FN}, \tag{2}$$

dosiahla hodnotu 66,66%. Špecifickosť, vypočítaná ako pomer správne klasifikovaných negatívnych príkladov oproti všetkým negatívnych príkladom

$$\frac{TN}{TN+FP}, \tag{3}$$

dosiahla hodnotu 85,4%.

Vyššie uvedené výsledky spolu s vygenerovanými pravidlami boli prezentované tímu, ktorý zodpovedá za kvalitu. Zistené pravidlá posúdili za zmysluplné, pričom ich najviac zaujalo, že ak bude vlastnosť suroviny X3 vyššia ako 54,25, tak bude výsledná kvalita na 80,95% vysoká.

IV. ZÁVER

Keďže v dnešnej dobe sa počas výroby generuje obrovské množstvo dát a aj malé vylepšenia môžu mať významný vplyv na výslednú kvalitu výrobkov, výrobné firmy javia veľký záujem o poznávanie vzorov v historických dátach na predpovedanie budúceho správania. Aj z tohto dôvodu sa dolovanie dát, rýchlo rastúci interdisciplinárny odbor, stáva čoraz viac dôležitým i vo výrobnej oblasti

V tejto práci sme sa zaoberali klasifikáciou výslednej kvality vyrobeného vlákna, s cieľom spoznať, ktoré vlastnosti majú vplyv na výslednú kvalitu vlákna. Využili sme pritom rozhodovacie stromy, ktoré sú obľúbené pre ich jednoduchú interpretáciu. Výsledky sme nakoniec prezentovali tímu, ktorý v spoločnosti zodpovedá za kvalitu. Ten zistené pravidlá posúdil za zmysluplné, pričom ich najviac zaujalo pravidlo týkajúce sa vlastnosti suroviny X3.

POĎAKOVANIE

Táto práca bola podporovaná KEGA grantom MŠVVaŠ SR č. 065TUKE-4/2011 (50%), ako aj projektom Rozvoj Centra informačných a komunikačných technológií pre znalostné systémy (kód ITMS projektu: 26220120030) na základe podpory operačného programu Výskum a vývoj financovaného z Európskeho fondu regionálneho rozvoja (50%).

LITERATÚRA

[1] V. Garcia, J.S. Sanchez, R.A. Molineda, On the effectiveness of preprocessing methods when dealing with different levels of class imbalance, Knowledge-Based Systems, 25 (1) (2012), pp. 13-21.
[2] T. Fawcett, An introduction to ROC analysis, Pattern Recognition Letters, 27 (8) (2006), pp. 861-874.
[3] G. Koksal, I. Batmaz, M. C. Testik, A review of data mining applications for quality improvement in manufacturing industry, Expert Systems with Applications, 38 (10) (2011), pp. 13448–13467.
[4] Maki H., Teranishi Y. "Development of automated data mining system for quality control in manufacturing" in *Data Warehousing and Knowledge Discovery*, Springer Berlin / Heidelberg, 2001, pp. 93-100.
[5] Bakır, B., Batmaz, I., Gunturkun, F. A., Ipekci, _I. A., Koksal, G., & Ozdemirel. ―Defect cause modeling with decision tree and regression analysis.‖ In *Proceedings of XVII international conference on computer and information science and engineering*, 8–10 December 2006. Cairo: World Enformatica Society (Vol. 16, pp. 266–269).
[6] The R Project for Statistical Computing [http://www.r-project.org/]
[7] Package 'discretization' [http://cran.r-project.org/web/packages/discretization/discretization.pdf]
[8] Chapman P., Clinton J., Kerber R., Khabaza T., Reinartz T. ―CRISP-DM 1.0 Step by step data mining guide‖ 2000 SPSS Inc.
[9] J. Paralič, Objavovanie znalostí v databázach, Košice (2003)
[10] A. Lukáčová, A review of data mining applications in manufacturing, 12th Scientific Conference of Young Researchers (SCYR 2012)
[11] R. Kerber, ChiMerge: Discretization of numeric attributes, AAAI, 199, pp. 123-128.
[12] K.B. Irani, J. Cheng, U.M. Fayyad, et al., Applying machine learning to semiconductor manufacturing, IEEE Expert 8 (1) (1993), pp. 41–47.
[13] Ciflikli C., Implementing a data mining solution for enhancing carpet manufacturing productivity. In Knowledge-based systems, 23 (2010), pp. 783-788.

Social Web Mapping

Kristína MACHOVÁ, Tomáš RAKUŠČINEC

Department of Cybernetics and Artificial Intelligence, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Slovak Republic

kristina.machova@tuke.sk, tomas.rakuscinec@student.tuke.sk

***Abstract* — The main focus of this paper is on dynamic analysis of social networks. The paper describes the results of the dynamic analysis of the data, which were collected from social network Internet Relay Chat (IRC). The network IRC has a character of discussion channel, which serves for solving user's problems.**

***Keywords* — social web, web mapping, social network analysis, discussion channel**

I. INTRODUCTION

The social web can be considered as an upgrade of the classic web. The classic web can be illustrated with an idea of a word-wide billboard – anybody can publish some information piece or make it accessible for public inspection on the billboard (anybody who has necessary skills in web page creation – but considerably greater amount of web users have abilities only for reading published information). On the other hand, the social web or web 2.0 reinforces social interactions among users and provides an opportunity for great majority of web users to contribute to web content. It can be said, that it increases the number of web content providers. Social interactions among users are enabled by communication within social networks, by the possibility to contribute to web discussions, and so on. Our work focuses on social networking.

The concept "social network" is frequently used and it represents one important form of social on-line life of many people. Modern social networks help users to maintain connections and contacts to other users without dependency on a geographical locality. The social network cannot be mapped in its entirety. Mapping of some of well known social network (for example Facebook) is a difficult task, when we want to visualize it in a way, which guarantees a readability of this visualization. So, our work concentrates on small networks and so as on network dynamics visualization.

II. SOCIAL NETWORKS

Social network is a set of entities (nodes, actors), which are connected. This connection can be represented by one or more types of relationships. According [5] social network is heterogenic and multi-relational data set represented by graph. Nodes and edges of this graph have attributes and objects (nodes) can be classified into categories. The relations can be oriented and need not be binary only. Social networks can be considered not only in social, but also in technological, economic or biologic context. Examples of such networks can be electrical distribution network, telecommunication networks, (computers) viruses spreading, collaborative networks (co-authors), citing networks, web networks and so on. According [9] the main concepts in social networks analysis are: *actor, relation, dyad, triad, subgroup, group and network*.

An *actor* is social entity in a network. The aim of social network analysis is to make sense of relations between these entities and to analyses consequences of these relations. Actor can be discrete individual, organization or simply group of social entities. The actors represent basic units of social network in the process of its analysis.

Individual actors are connected each other by social ties. Such social tie can be: friendship, respect, preference, association, affiliation, interaction and so on. Collection of social ties between actors (entities) is titled *relation*. The relation $R_v \subseteq \{(A \times A) \times R\}$ between actors of the same type is binary relation, where A is the set of actors, R is the set of relations and v_x is intensity of relation.

The tie between two actors is called *dyad*. A *triad* arises by extension of the dyad by one more actor. A *subgroup* is extension of the triad. The subgroups play important role in the social networks dynamic analysis. The network dynamics represents moving of actors within the

social network that is formation and expiration of subgroups. A group is such set of actors, which belong to common and finite set and which are connected with measurable ties. So *network N* can be defined as pair of the set of actors *S* and the set of all relations *R* over the set of actors: $N=(A, R);\ A=\{A_1,\ldots,A_X\};\ R=\{R_1,\ldots,R_Y\}$.

One of objectives of social networks analysis is identification of cohesive sub-groups and their analysis. The cohesive sub-groups are sub-sets of actors, there are strong, direct, intensive and frequent connections among them. They can be identified on the basis of a number and complexity of mutual connections, on the basis of closeness or readability of other actors of given group and on the basis of connections frequency between actors.

A. Social Network Types

The social networks can be divided into types according various points of view. For example according number and type of edges (directed, undirected networks, networks with cycles and networks with multi-edges), according way of edges evaluation (weighted, un-weighted, marked and temporal networks), according number of actor types (monoecious networks – contains only one type of actors, dioecious networks – contains more types of actors) and according number of relation types (one-relation networks, multi-relation networks). During real networks modelling various network types can be distinguished: small-word phenomenon networks, scale-free networks, social cycle networks, random and lattice networks.

Real networks are networks, which rise in process of self-organization. This self-organization process can optimize: conservation of a network local structure, conservation of a good communication between network nodes and good resistance against random errors. An example of the real network is Internet or network of professional contacts or telecommunication networks.

The great majority of nodes of the *small-word phenomenon networks* [6] are not connected to other nodes directly, but they are achieved from other nodes via small number of edges. The small-word networks have concise shortest paths, high clustering coefficient and small node separation (mentioned structural characteristics will be described in the next section). Interesting property of these networks is tendency to contain cliques and n-cliques (because of high clustering coefficient) and many nodes with high degree (see next section). Great number of nodes with high degree can lead to high distribution. The *scale free networks* have distribution of a degree minimally depended on Power Law. More details can be found in [1]. The *random and lattice networks* have clustering coefficient close to zero. Its node separation is grown slowly with node number growing [7]. The *social cycle network* model represents networks with feedback [3]. Active nodes (actors) of the feedback network communicate through the network to coordinate the process of joining of new actors and to create new relations. For realization of this process, three parameters are needed: level of actor activity, distance shorting (how soon the actor fails in the effort to find suitable partner for new relation) and range of searching within the network.

B. Structural Characteristics of Networks

Some properties of some social networks can be researched with the aid of its structural characteristics. The most important and useful characteristics are: node degree, shortest paths and clustering coefficient [2].

Node degree k represents total number of its connections. In physical literature, this characteristic is usually called "connectivity", which has different meaning in graph theory. The node degree consist of "in-degree" and "out-degree" $k=k_i+k_o$. The in-degree is the number of incoming edges of a node and out-degree is the number of outgoing edges. It holds:

$$P(k)=\sum_{k_i}P(k_i,k-k_i)=\sum_{k_0}P(k-k_0,k_0), \qquad (1)$$

where: $P(k)$ is the degree distribution and $P(k_i,k_0)$ is the joint in-degree and out-degree distribution. The in-degree distribution $P_i(k_i)$ and the out-degree distribution $P_0(k_0)$ (more obvious are notations $P(k_i)$ and $P(k_0)$) can be stated according formula (2):

$$P_i(k_i)=\sum_{k_0}P(k_i,k_0), P_0(k_0)=\sum_{k_i}P(k_i,k_0) \qquad (2)$$

If given network has no connections to exterior space, then the average in-degree is equal to the average out-degree (see formula (3)).

$$\overline{k_i} = \sum_{k_i,k_0} k_i P(k_i,k_0) = \overline{k_0} = \sum_{k_i,k_0} k_0 P(k_i,k_0) \qquad (3)$$

The node degree is a local characteristic, but we shall see that a degree distribution can determines some important global characteristics of random networks. The average node degree within the whole network represents *network connectivity*.

Shortest path is defined as geodetic distance of two nodes *u* and *v*, which represents the shortest one from all possible paths between these two nodes l_{uv}. The shortest path l_{uv} need not be the same as l_{vu}. The average shortest path between all node pairs within the entire network is often called the diameter of a network. It is related to an average separation of pairs of nodes.

This *node separation* represents one another global property of a network – closeness of nodes. Node separation *l* is averaging shortest distance $d_{min}(a,b)$ of the nodes *a* and *b*.

$$l = \frac{1}{N}\sum_{i=1}^{N} d_{\min}(a,b) \quad (4)$$

Let us consider a network with undirected edges. Then the number of all possible connection of the nearest neighbours k_i of a node *u* is $\binom{k_i}{2} = \frac{k_i(k_i-1)}{2}$. Let only E_i are present. Then *clustering coefficient* of the node *u* can be counted according formula (5).

$$C_i = \frac{2E_i}{k_i(k_i-1)}. \quad (5)$$

C. Network Data

Social networks are sources of structural, composition and temporal data. The structural data are obtained from connections between actors (for example friendships between people, business transactions between firms and so on). The structural data can be represented by quadratic binary matrix (1 represent connection and 0 represent absence of connection). The composition data introduce attributes of network actors, which can be direct (for example interests, age, nation, country and so on) and indirectly obtained in the process of network analysis by counting (for example node degree, node centrality and so on). These data are static and they can be represented by orthogonal matrix. Sometimes we are interested in dynamics of these data, which can be research in the process of dynamic analysis of social networks. This dynamic analysis uses temporal data, which describe dynamic properties of social network. The temporal data represent some time change and its time evidence.

III. Dynamics of Discussion Channel

During dynamic analysis of network, our attention was focused on social network IRC (Internet Relay Chat), which was created on the basis of interactions between users of discussion channel located on web. The advantage of this selection is, that users of this discussion channel continuously communicate between each other long time, because of themes of their communication is always actual for them. These themes are related to technical problem solving, to technical news discussion or to common communication. The IRC social network is created continuously in the process of communication between users per running time. So it is naturally dynamic system suitable for dynamics analysis. The dynamics of this system is created by emergency. The dynamics of discussion channel is not created purposelessly by user but it is related to actual load of this channel according to real need to solve some technical problems, which have arisen.

Our tests were from the beginning concentrated on percentage of load of researched discussion channel during particular phases. Each phase is represented by one hour of a day. Figure 1 illustrates percentage of average loading of discussion channel in particular hours of the day.

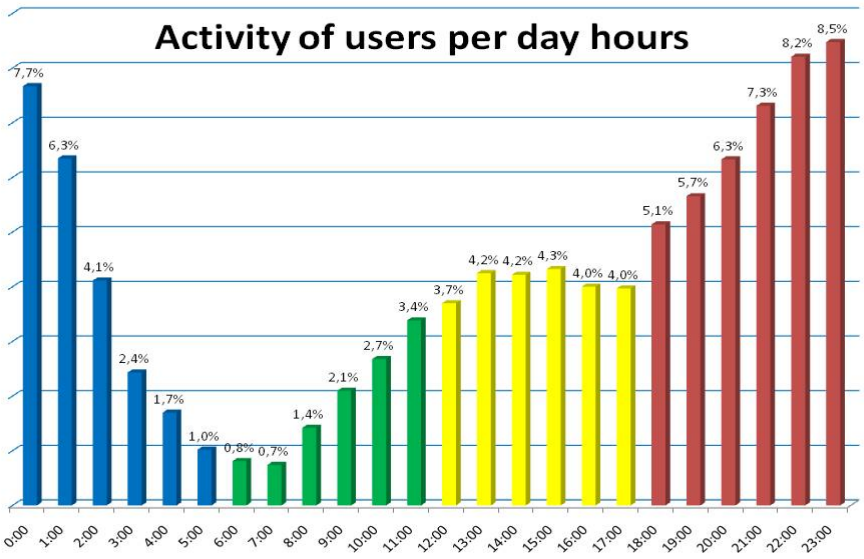


Fig. 1 Average percentage of discussion channel IRC loading. Each column represent one hour of the day. Data were collected during 211 days.

The whole number of users' communications within all 211 days was 1672263 iterations. Activities on the discussion channel were observed with the aid of autonomous robot Eggdrop [4]. The robot fulfils a role of manager of users. It provides some added functionalities for operators, moderators and functionalities for safety measures.

IV. Conclusion

The purpose of this work was a research of dynamics of the IRC social network. We have analysed data, which were collected by PieSpy [8]. The whole number of collected screens was 181310. It takes 5 GB of disc space. The social network was monitored during enough long time (since October 9, 2009 to December 05, 2010) to trace permutations in the social network organization. Dynamic processes originate in this social network by emergency. It means that network users do not communicate with other users for purpose of social network building or for purpose of connections gaining, but because of technical problem solving. This social network is a result of independent communication on public level.

Acknowledgment

This work is the result of the project implementation Development of the Centre of Information and Communication Technologies for Knowledge Systems (project number: 26220120030) supported by the Research & Development Operational Program funded by the ERDF (100%).

References

[1] Barabasi, A.L., "Scale-free Network". Scientific American, 288, 2003, p. 60-69.
[2] Dorogovtsev, S.N., Mendes, J.F.F., "*Evolution of Networks." Departamento de Fisica and Centro de Fisica do Porto, Faculdade de Ciencias, Universidade do Porto, Portugal, 2001.*
[3] Douglas, R. *et al., "Generative Model for Feedback Networks." In Physical Review E, 2006.*
[4] EGGDROP *The word's most popular open source Internet Relay chat bot.* [cit. 2011-04-11]. Available at: http://www.omgirc.com/index.php?itemid=26.
[5] Hann, W. P., "*Hyperlink Network Analysis: A New Method for the Study of Social Structure" on the Web Royal Netherlands Academy of Arts & Sciences,* Amsterdam, The Netherland, CONNECTIONS, 2003, Vol. 25, No 1, p. 49-61.
[6] Kleinberg, J., "The Small-Word Phenomenon: An Algorithmic Perspective". Proc. of the 32nd ACM Symposium on Theory of Computing, Portland, Oregon, US, 2000, p. 163-170.
[7] Markošová M. "Networks Dynamics." In: Kvasnička et al. (ed.) Artificial Intelligence and Cognitive Sciences II. Bratislava (SK): STU, 2010, p. 321-377.
[8] PIESPY *PieSpy Social Network Bot.* [cit. 2011-04-14]. Available at: http://www.jibble.org/piespy/.
[9] Repka M., "*Analysis of selected types of social network".* Technical University in Košice, 2011. 75 pp.

Návrh a realizácia riadenia križovatky PLC automatom

Gergő MARCI, Matej ČOPÍK, Ján SARNOVSKÝ

Katedra kybernetiky a umelej inteligencie, Fakulta elektrotechniky a informatiky,
Technická univerzita v Košiciach, Slovenská republika

gergo.marci@student.tuke.sk, matej.copik@tuke.sk, jan.sarnovsky@tuke.sk

***Abstrakt* — Článok sa venuje návrhu a realizácie riadenia modelu viacúrovňovej križovatky PLC automatom. Model je k PLC automatu pripojený cez jednotku vzdialených vstupov a výstupov pripojenej k PLC automatu cez ethernet rozhranie. Na model križovatky je vytvorená aj vizualizácia v prostredí Intouch od spoločnosti Wonderware.**

***Kľúčové slová* — križovatka, PLC automat, vizualizácia**

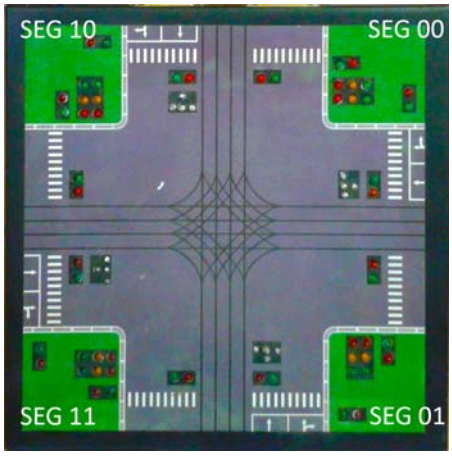
I. ÚVOD

S rozšírením cestnej premávky svetelné signalizačné zariadenie na križovatkách sa stali nevyhnutnou časťou riadenia dopravy. V dnešnej dobe sa skoro na každej križovatke stretávame so svetelnými signalizačnými zariadeniami, ktoré popri obyčajnom riadení premávky sú väčšinou navrhnuté tak, aby sa snažili minimalizovať čakaciu dobu na križovatke s ohľadom na využiteľnosť ciest v križovatke.

II. MODEL KRIŽOVATKY

A. Križovatka

Model viacúrovňovej križovatky je ovládaný 16-timi výstupmi. Vzhľadom k pomerne značnému počtu LED diód je križovatka rozdelená na štyri segmenty. Každý segment obsahuje 18 LED diód. Párové signalizačné svetlá určené pre chodcov sú zapojené paralelne. Jeden segment je ovládaný 14-timi riadiacimi signálmi. Posledné dva signály sú vyhradené na prepínanie segmentov. Jednotlivé segmenty sú prepínané vysokou rýchlosťou a práve to robí dojem že segmenty svietia naraz a križovatku vnímame ako jeden celok. Výber segmentu riadi BCD prevodník spolu so spínacími tranzistormi. Jednotlivé segment sú znázornené na obrázku č.1.[1]



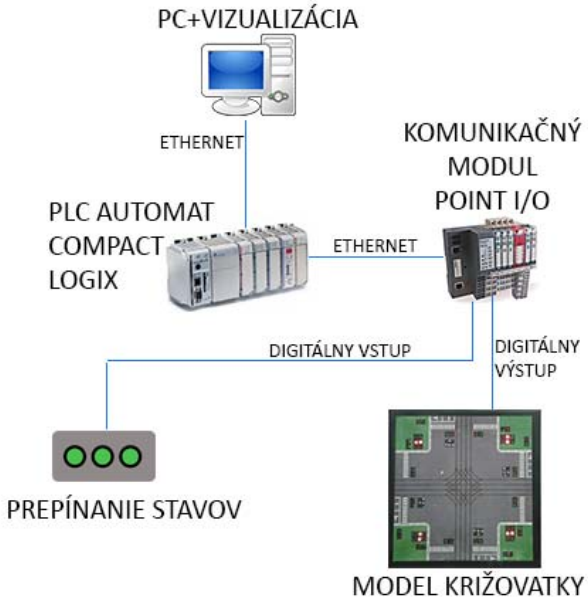
Obr.1. Model križovatky s vyznačenými segmentmi

B. PLC automat

Na riadenie je používaný PLC automat CompactLogix 1769-L23E-QBFC1B od firmy Allen-Bradley. PLC automat má zabudované nasledujúce moduly:

- digitálny vstupný modul pre 16 vstupov
- digitálny výstupný modul pre 16 výstupov
- analógový vstupný modul pre 4 vstupy
- analógový výstupný modul pre 2 výstupy
- modul HSC(High Speed Counter)[3]

PLC automat je následne pripojený ku vzdialenému komunikačnému modulu Point I/O 1734-FPD taktiež od firmy Allen-Bradley. V module sú zapojené dva výstupné karty Point I/O Output Module 1734-OB8 a jedna vstupná karta Piont I/O Input Module 1734-IB4. Komunikácia medzi PLC automatom a vzdialeným komunikačným modulom je zabezpečená sieťou Ethernet. Štruktúra modelu s ovládacími prvkami je znázornená na obrázku č.2.



Obr.2. Štruktúra modelu križovatky

III. RIADENIE KRIŽOVATKY

A. Riadiaci program

Na programovanie PLC automatov typu CompactLogix slúži aplikačné programovacie rozhranie RSLogix 5000. Program je vyhotovený formou rebríkových schém. Rebríkové schémy obsahujú tzv. "rung"-y, ktoré reprezentujú pravidlá. Pravidlá sú sekvenčne vykonávané v jednej nekonečnej slučke.

Hlavný program obsahuje sekvenčné prepínanie jednotlivých segmentov so vzorkovaciou frekvenciou 250 Hz. Po spustení programu sa spustí časovač o dĺžke 4 ms. Po ukončení časovania inštrukcia SQO v cykle aktivuje postupne aktivuje segment-urobí kombinácie z posledných dvoch bitov- a cez masku pošle na druhú výstupnú kartu. Po ukončení úlohy SQO inštrukcie časovač sa vynuluje a znova sa spustí. To je základné riadenie križovatky a opakuje sa kým to užívateľ nezastaví. V hlavnom programe sú na základe 3 tlačidiel zavolané podprogramy pre jednotlivé režimy fungovania križovatky. V podprogramoch sa prepínajú stavy cyklicky s daným časovým intervalom. Rozsvecovanie LED diód sa vykonáva aplikovaním príslušnej masky – maska pre 8 výstupov na prvej výstupnej karte a pre 6 výstupov na druhej výstupnej karte. Posledné 2 bity na druhej výstupnej karte sú vyhradené na prepínanie segmentov.

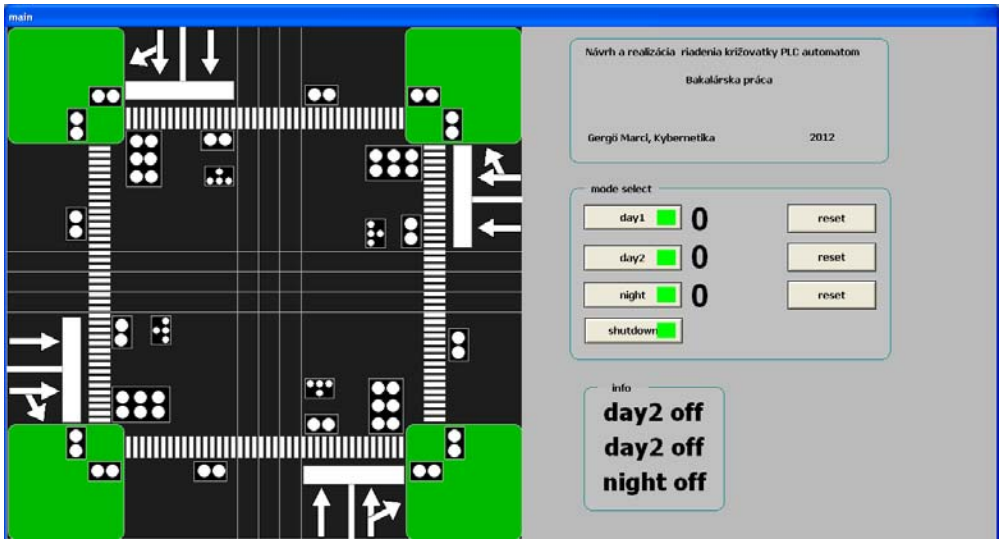
Spojenie medzi PLC automatom a počítačom je zabezpečené pomocou komunikačného servera RSLinx od firmy Rockwell Automation.

B. Vizualizácia

Na riadenie modelu križovatky pomocou vizualizačného systému bol použitý vizualizačný systém InTouch od firmy Wonderware. Je to systém pre vizualizáciu, zber dát a supervízne riadenie technologických procesov kategórie SCADA / HMI (Supervisory Control and Data Acquisition / Human-Machine Interface) s celosvetovým rozšírením. Umožňuje ľahko vytvoriť grafické zobrazenie akýchkoľvek výrobných technológií na monitore počítača, ich ovládanie a dynamické animácie, ktoré sa znázorňujú v reálnom čase a zobrazujú aktuálne stavy prevádzkovaných technologických systémov vo výrobe. InTouch slúži aj ako jednotné vizualizačné rozhranie (front-end) pre ďalšie produkty od firmy Wonderware.[2]

Spojenie medzi vizualizáciou a PLC automatom je zabezpečené pomocou komunikačného servera ABCIP DAServer od firmy Wonderware.

Pri realizácii vizualizácie sa vyskytol problém, čo značne ovplyvnil realizáciu. Prenos dát medzi vizualizácie a PLC automatom cez DAServer je veľmi pomalý a kvôli nedostatku údajov by vizualizácia nedávala reály obraz o fungovaní križovatky, preto bolo navrhnuté alternatívne riešenie. Cyklické prepínanie segmentov sa realizuje lokálne vo vizualizácii pomocou skriptov a do vizualizácie je prenesený len čas potrebný na prepínanie stavov križovatky a riadiace signály na výber režimu.



Obr.3. Vizualizácia prostredníctvom systému InTouch

IV. ZÁVER

Aj napriek problémov s vizualizáciou bol vytvorený plne funkčný riadiaci systém križovatky. V budúcnosti by bolo vhodné rozšíriť model križovatky o kamerový systém na sledovanie vozidiel a tak efektívnejšie riadiť premávku so znížením čakacej doby pri slabej premávke.

Model disponuje snímačom na simulovanie detekcie pohybu.V reálnych aplikáciach túto úlohy plní kamerový systém doplnený o softvér na detekciu prítomnosti vozidiel. Ak kamerový systém zaznamená prítomosť vozidla zapne riadiaci systém svetelnej signalizácie.

POĎAKOVANIE

Táto práca bola podporená z vedeckého projektu Vega č. 1/0286/11 Grantovej agentúry SR pod názvom „Dynamické hybridné architektúry v multiagentových sieťových riadiacich systémoch“.

LITERATÚRA

[1] KAČMÁR, Matej: Model viacúrovňovej križovatky: Bakalárska práca. Košice: TU FEI, 2009. 62 s.
[2] http://pantek.cz/produkty/wonderware-software/
[3] Rockwell Automation Inc.: CompactLogix Controllers Specifications [online], 2012. Dostupné na internete:< http://literature.rockwellautomation.com/idc/groups/literature/documents/td/1769-td005_-en-p.pdf>

Využitie vhodných webových technológií pre prezentáciu a rezerváciu ubytovacích kapacít

Tomáš ORSÁG, František BABIČ

Katedra kybernetiky a umelej inteligencie, Fakulta elektrotechniky a informatiky,
Technická Univerzita v Košiciach, Slovenská republika

tomas.orsag@student.tuke.sk, frantisek.babic@tuke.sk

***Abstrakt* — Ubytovací portál alebo server poskytuje možnosť ako si jednoducho a rýchlo nájsť a rezervovať ubytovanie podľa predstáv návštevníkov. V tomto prostredí existuje viacero alternatív, ktoré rôznymi spôsobmi poskytujú svoje služby. Na základe analýzy konkurenčných riešení bol vytvorený zoznam požiadaviek, ktoré bude nová verzia portálu www.spanie.sk spĺňať, aby sa zaradila medzi najnavštevovanejšie a najvyužívanejšie ubytovacie portály na Slovensku. Dôležitý faktor v tomto procese predstavovala vhodne implementované SEO optimalizácia, ktorá viedla k podstatnému zvýšeniu návštevnosti portálu.**

***Kľúčové slová* — ubytovací portál, SEO optimalizácia**

I. Úvod

Stále výraznejší vplyv informačno-komunikačných technológií ovplyvnil aj oblasť cestovného ruchu. Vznik ubytovacích portálov znamenal prevrat vo vyhľadávaní a rezervovaní ubytovania na celom svete. On-line rezervačné systémy nahradili rezervačné knihy, a tým uľahčili na jednej strane prácu ubytovateľom a na druhej strane proces rezervácie potenciálnym hosťom. Výnimkou nie je ani cestovný ruch na Slovensku. Za posledných desať rokov rapídne vzrástol počet ubytovacích serverov na Slovensku, ktoré poskytujú široký sortiment služieb pre poskytovateľov ubytovacích služieb. Pre ľudí ponúkajú rýchly a jednoduchý spôsob ako vyhľadať a rezervovať ubytovanie v konkrétnej lokalite. Ubytovacie portály fungujú buď na princípe cestovej kancelárie (ubytovací server vystupuje ako sprostredkovateľ medzi klientom a ubytovacím zariadením) alebo na princípe priameho kontaktu na ubytovateľa (v tomto prípade je v internetovej prezentácii konkrétneho zariadenia zverejnený aj telefonický kontakt, e-mailová adresa, prípadne odkaz na vlastnú webovú stránku zariadenia).

Tento trh predstavuje rýchlo rastúce prostredie, do ktorého patrí aj webový portál www.spanie.sk. Pôvodná verzia tohto portálu už nespĺňala požiadavky kladené na kreatívne a užívateľsky prívetivé webové stránky, preto tento fakt motivoval vznik projektu, ktorého výsledkom bude nová verzia portálu založená na vhodných webových technológiách a kvalitnej SEO optimalizácii.

Tento článok je rozdelený do troch hlavných častí, kde prvá popisuje motiváciu a vybrané konkurenčné riešenia; druhá časť je venovaná procesu vývoja a implementácie novej verzie portálu; a posledná časť vyhodnocuje vytvorené riešenie z pohľadu návštevnosti, počtu ubytovacích zariadení a uskutočnených rezervácií.

A. Pôvodná verzia portálu

Portál www.spanie.sk je v prevádzke od roku 2004, keď vznikol ako prezentácia ubytovacích zariadení v Slovenskom raji, nachádzajúcich sa v okolí zariadenia majiteľa stránky. Cieľom bolo zlepšiť informovanosť o možnostiach cestovného ruchu v tejto lokalite pomocou aktuálne najdôležitejšieho komunikačného a informačného média – webu. Záujem o možnosť prezentovať svoje zariadenie zo strany poskytovateľov postupne rástol, čo viedlo k expanzii z ponuky ubytovania v obci Mlynky na ponuku zariadení na celom Slovensku. V priebehu pár rokov svojej existencie patril portál k najlepších na Slovenku, avšak postupom času sa záujmy pôvodného majiteľa preorientovali aj na iné obchodné činnosti, čo malo za následok, pri neustále sa zväčšujúcej konkurencii, úpadok servera. Z technického hľadiska pôvodná verzia používala zastaraný redakčný systém, ktorý pri stredne náročných zásahoch do portálu vyžadoval programátorské znalosti; chýbali jazykové mutácie, privátna zóna pre

poskytovateľov ubytovania s možnosťou pridávať a editovať informácie; a takisto SEO optimalizácia pre vyhľadávače bola na veľmi na nízkej úrovni, čo prispelo ku klesajúcej návštevnosti portálu.

B. Konkurenčné portály

V tejto časti budú popísané vybrané konkurenčné riešenia ubytovacích portálov na Slovensku. Tento krátky prehľad poukáže na základné funkcie, ktoré sú v rámci týchto portálov k dispozícii.

Limba.sk predstavuje ubytovací portál, ktorý funguje na princípe cestovnej kancelárie. Prostredníctvom tohto portálu si každoročne rezervuje ubytovanie viac ako 150 000 klientov z celého sveta. Propaguje všetky typy zariadení ako chaty a chalupy, penzióny, hotely a ďalšie. Limba.sk okrem dovoleniek organizuje aj kongresy, školenia, relaxačné a kultúrne podujatia. Vyhľadávanie je rozdelené na základe jednotlivých svetadielov, ktoré sú podrobnejšie rozčlenené na štáty. V rámci Slovenskej republiky momentálne portál ponúka 3327 ubytovaní. Prezentácia ubytovacieho zariadenia obsahuje popis zariadenia, rozloženie a typy izieb, celkovú kapacitu, fotogalériu, cenovú kalkuláciu, vybavenosť objektu a jednotlivých izieb, kalendár obsadenosti a turistické, kultúrne a športové objekty v okolí. Rezerváciu ubytovania je možné zrealizovať prostredníctvom internetového rezervačného formuláru alebo telefonicky. V rezervácii je nevyhnutné uviesť presný termín príchodu, počet osôb a zoznam vybraných objektov prioritne zoradených. Pre záväznú rezerváciu je potrebné zaplatiť preddavok do 2 dní, pričom výška sa líši v závislosti od konkrétneho ubytovania. (oficiálne informácie)

Projekt *Ubytovanienaslovensku.sk* bol spustený v roku 2007 ako ubytovací portál, ktorý funguje na princípe priameho kontaktu na ubytovateľa pomocou údajov umiestnených priamo v prezentácii príslušného zariadenia. Návštevníci môžu takisto využiť možnosť on-line rezervácie, ktorá príde následne priamo na e-mail ubytovateľa. Klienti tak majú možnosť získať okamžite presné a aktuálne informácie ohľadom obsadenosti kapacít a cenách za ubytovanie z prvej ruky. Takýto spôsob vyhľadávania zaručuje najnižšie ceny, nakoľko nie je zaťažený poplatkami za sprostredkovanie ubytovania cestovnými kanceláriami. (oficiálne informácie)

II. NÁVRH A VÝVOJ

A. Návrh riešenia

Na základe analýzy konkurenčných portálov a doterajších skúsenosti s poskytovateľmi ubytovania bol vytvorený zoznam požiadaviek, ktoré by mala nová verzia portálu spĺňať, aby ponúkal všetky potrebné funkcie v užívateľsky príjemnom prostredí.

Kvalitný a prepracovaný redakčný systém je základom každého moderného portálu. Je potrebné, aby umožnil správu portálu aj bez znalosti programovacích jazykov. V tomto prípade by umožnil správu štruktúry portálu, možnosť pridávať akciové ponuky, možnosť zasielanie informačných letákov, kompletnú správu ubytovacích zariadení, fakturačný systém a správu fotogalérií. Dostupné jazykové mutácie umožnia získať návštevníkov aj mimo Slovenska, ktorý budú schopní vyhľadať a rezervovať ubytovanie. Vďaka internému rozhraniu pre ubytovacie zariadenia získajú poskytovatelia možnosť editovať informácie o svojom zariadení v reálnom čase, čo zároveň zníži zaťaženosť samotných pracovníkov portálu v oblasti správy obsahu.

Ako nevýhoda pri väčšine konkurenčných riešení bola identifikovaná prezentácia jednotlivých ubytovacích zariadení ako celok, nie ako jednotky (napr. izba, apartmán, atď.). Takýmto spôsobom funguje napr. známy medzinárodný ubytovací portál www.booking.com. Práve touto novinkou na Slovenskom trhu by mala nová verzia portálu osloviť nielen majiteľov ubytovacích zariadení, ale aj návštevníkov, ktorí ubytovanie hľadajú.

Nevyhnutnou súčasťou internetových stránok je SEO optimalizácia pre vyhľadávače. Nakoľko portál je zameraný na vysoko konkurenčné kľúčové slová, kvalitná on-page optimalizácia nebude postačujúca na dobrú pozíciu vo vyhľadávačoch, a z tohto dôvodu bude potrebné pracovať aj na veľmi dobrej off-page optimalizácii.

B. Návrh GUI

Úvodná stránka portálu pozostáva z dvojstĺpcového layout-u, viď. Obr.1. Šírka webovej stránky bola nastavená na 980 pixelov z dôvodu optimalizácie aj pre rozlíšenie 1024x768.



Obr. 1 Nová verzia portálu www.spanie.sk

C. Použité technológie

Na vytvorenie novej verzie portálu boli použité nasledujúce technológie: MySQL ako databázové riešenie pre uloženie dát; XHTML na vytvorenie základnej štruktúry portálu; CSS na zadefinovanie vzhľadu základnej štruktúry portálu; objektovo orientované PHP; Google Maps API pre generovanie mapy; a ďalšie techniky na správu webu ako JavaScript; AJAX pre vývoj interaktívnych webových aplikácií, ktoré umožňujú meniť obsah stránok bez potreby ich kompletného znovu načítania zo serveru; jQuery knižnica, ktorá umožňuje jednoduchšiu navigáciu dokumentu, výber DOM elementov, vytváranie animácií, spracovanie udalostí, a vývoj AJAX aplikácií.

V rámci on-page SEO optimalizácie boli zapracované používateľsky prívetivé linky, upravené „title tagy“, pridané nadpisy a štruktúrovane zoznamy, atď. Všetky hlavné on-page prvky sú generované automaticky pre každú podstránku na základe určených kritérií v systéme. Off-page optimalizácia bola prevažné zameraná na hľadanie kvalitných a korektných spätných odkazov. Okrem toho boli využité metódy „social bookmarking“ a propagácia prostredníctvom vytvorených článkov a webových prezentácií.

III. ZÁVER

Tento článok popisuje spôsob ako využiť vhodné webové technológie pre prezentáciu ubytovacích zariadení v konkurenčnom prostredí na Slovensku. V rámci tohto prostredia existuje niekoľko alternatív, ktoré ponúkajú viac či menej podobné funkcie, preto je dôležité ponúknuť návštevníkom niečo nové, čím sa daný portál odlíši od konkurencie a umožní rezerváciu ubytovania používateľsky príjemným spôsobom. Výsledná verzia portálu www.spanie.sk v sebe zahŕňa výhody iných portálov a snaží sa eliminovať možné nevýhody. . Hlavnou výhodou je prezentácia jednotlivých ubytovacích jednotiek a nie ubytovacieho zariadenia ako celku, čím je možné poskytnúť záujemcom o ubytovanie kompletné informácie o zariadení a majiteľom ubytovacích zariadení korektné informácie o jednotlivých objednávkach. Tento prístup ponúka aj napr. jeden z najnavštevovanejších ubytovacích a rezervačných portálov na svete – www.booking.com.

Priemerná návštevnosť portálu bola v období november 2010 po september 2011 okolo 1250 návštev za mesiac. V októbri 2011 bola spustená nová stránka len so základnou SEO optimalizáciou, čo neprinieslo žiadne výrazné zmeny v priemernej návštevnosti do konca roka 2011. Po zapracovaní SEO optimalizácie a realizovanej marketingovej kampani priemerná návštevnosť v období prvých štyroch mesiacov 2012 stúpla na vyše 16 000 návštev za mesiac, čo je zvýšenie priemernej návštevnosti o 14 750 návštev/ mesiac. Takisto nasledovalo zvýšenie počtu registrovaných ubytovacích zariadení o 50% a zvýšenie počtu rezervácii ubytovania zo strany návštevníkov portálu o 45%. Úspešnosť nasadenia funkcie prezentácie ubytovacích zariadení po jednotkách dokumentuje anketa na úvodnej stránke, viď. Obr.1.

POĎAKOVANIE

Tento príspevok vznikol za podpory KEGA projektu č. 065TUKE-4/2011 (50%) a realizáciou projektu Rozvoj Centra informačných a komunikačných technológií pre znalostné systémy (kód ITMS projektu: 26220120030) na základe podpory operačného programu Výskum a vývoj financovaného z Európskeho fondu regionálneho rozvoja (50%).

Informačný systém pre online prieskumy spokojnosti

Beáta PETRÍKOVÁ, Iveta ZOLOTOVÁ

Katedra kybernetiky a umelej inteligencie, Fakulta elektrotechniky a informatiky
Technickej univerzity v Košiciach, Slovenská republika

beata.petrikova@student.tuke.sk, iveta.zolotova@tuke.sk

Abstrakt — Článok pojednáva o možnostiach tvorby online prieskumov spokojnosti prostredníctvom dostupných webových aplikácií, platených či voľne dostupných, sumarizuje ich výhody a nevýhody s dôrazom na produkt SurveyMonkey. Prezentuje aj praktickú realizáciu online prieskumu, návrh a implementáciu vlastného informačného systému postaveného na IBM Domino technológii.

***Kľúčové slová* — dotazník, informačný systém, online prieskum**

I. Úvod

Čo je to prieskum? Prieskum nie je len systém pre zber informácií. Často sa v diskusiách o vykonávaní prieskumov kladie dôraz iba na nesprávne dotazníky. Ak sa ale majú prieskumy využívať čo najefektívnejšie, je dôležité si uvedomiť, že dotazník je len jedným z prvkov procesu, ktorý začína definovaním cieľov a končí analýzou dát a reportovaním výsledkov.

A. Základné informácie ku prieskumom

Základné kroky v procese prieskumu:

- **Definovanie cieľov** (Určiť, čo chcete vedieť a prečo, Premýšľať o tom, komu budú predkladané výsledky, Zvážiť vplyv externých zdrojov financovania),
- **Definovanie populácie a výber rámca odberu vzoriek** (Pozrieť sa na existujúce segmenty výberu alebo vytvoriť rámec pre výber vzoriek, Zvážiť pravdepodobnosť a nepravdepodobnosť stratégií odberu vzoriek),
- **Dizajn a stratégia zberu dát** (Zhodnotiť čas a rozpočtové obmedzenia, Odhadnúť dostupné zdroje, Vybrať si na prieskum správnu metódu)
- **Rozvíjanie dotazníka** (Napísať otázky, Vykonať predbežné testy fungovania a vyplnenia dotazníka),
- **Zber dát** (Monitorovať odpovede, Sledovať a upravovať zber podľa potreby)
- **Správa dát** (Vytvoriť číselníky, Pripraviť vstupné dáta, resp. ich importovať, exportovať, Vyčistiť dáta, Transformovať dáta do publikovateľnej a spracovateľnej podoby),
- **Analýza dát** (Spracovať dáta analytickými nástrojmi, verifikovať ich hodnovernosť, smerodajnú odchýlku, pravdepodobnosť),
- **Šírenie výsledkov** (Napísať správu, Vykonávať osobné prezentácie),
- **Vytvorenie vizuálnych pomôcok** (Šíriť poznatky a prípadne realizovať nápravné opatrenia).

Tradičné metódy pre správu prieskumov sú postavené na báze telefonického rozhovoru, rozposlaní dotazníkov mailom a osobného (face-to-face) dopytovania. Do týchto metód sa pridáva celý rad nových techník. K dispozícii sú možnosti internetu, najmä e-mail a prieskumy postavené na dostupnosti web technológii (web-based). V e-mailových prieskumoch je dotazník obsiahnutý v tele e-mailu správy alebo je v prílohe. Vo väčšine prípadov môže respondent vyplniť dotazník, ktorý zašle ako odpoveď na pôvodný e-mail.. V on-line prieskumoch je dotazník umiestnený na web stránke. Respondent navštívi miesto s web dotazníkom a to buď kliknutím na URL (Uniform Resource Locator) odkaz, ktorý sa nachádza v e-maily, alebo zadaním webovej adresy do okna prehliadača a dokončí vyplnenie dotazníka.

Ako u každého nového nástroja internetové prieskumy poskytujú mnoho príležitostí a výzvy pre ich používateľov. Je dôležité, aby si používatelia online prieskumov tieto možnosti a obmedzenia uvedomovali. Môžu tak prijímať relevantné rozhodnutia pri výbere metódy a

zberu dát. Je nepravdepodobné, že prieskumy postavené na dostupnosti internetu (internet-based) nahradia tradičné metódy zberu dát. Tieto metódy budú pravdepodobne ich alternatívou, resp. doplnkom [3].

B. Výhody a nevýhody online prieskumov

1) Výhody

- **Rýchlosť** – umiestnenie dotazníka na internet má potenciál zhromaždiť tisíce odpovedí počas niekoľkých hodín,
- **Cieľová skupina** – možnosť pridania odkazov na mnohých internetových stránkach s povolením a spoluprácou vlastníka tejto stránky,
- **Ekonomika** – internetové prieskumy sú najekonomickejší prostriedok, pomocou ktorého sa zhromažďujú údaje od veľkého počtu respondentov, ktorí sú geograficky rozptýlení.
- **Pridanie obsahu** – veľký počet internetových prieskumov a softvérových balíkov ponúka možnosť vkladania obrázkov, audio a video súborov,
- **Rozšírené typy otázok** – internetová stránka môže na vytvorenie dotazníka obsahovať širokú škálu otázok, podmienok,
- **Schopnosť klásť citlivé otázky** – nie je prítomný žiaden zadávateľ prieskumu a respondent môže vypĺňať dotazník vlastným tempom,
- **Anonymita** – nie je tu žiadna e-mailová adresa v súvislosti s internetovým prieskumom.

2) Nevýhody

- **Obmedzená populácia** – používanie internetu je zatiaľ stále doménou určitých populačných ročníkov,
- **Nekončenie prieskumu -** respondent môže ľahko skončiť s vypĺňaním na ľubovoľnom mieste.
- **Závislosť na softvéry** – internetové prieskumy si vyžadujú, aby používatelia používali softvér na vytvorenie a nasadenie dotazníkov. V ponuke na internete je veľa možností a balíkov pre online prieskumy. Produkty sa líšia cenou, jednoduchosťou použitia a flexibilitou. Zaujímavou možnosťou je použitie prieskumov formou SaaS (softvér ako služba).

C. Aplikácie na poskytovanie online prieskumov

Prieskum pomocou webových aplikácií robí proces prieskumu rýchlejším a efektívnejším ako predtým v reálnom čase, pretože prístup a výsledky sú k dispozícii 24 hodín denne. Dostupnými webovými aplikáciami sú napr.:

a) Zoomerang
b) FluidSurveys
c) MySurveyLab
d) SurveyMonkey
e) Survd
f) ZohoChallenge
g) Obsurvey
h) Rational Survey
i) Mineful
j) SISsurvey
k) Twiigs
l) PollAuthority
m) InfoPoll
n) Hostedsurvey

V tabuľke (Tabuľka 1) sú porovnané vybrané web aplikácie [1] s dôrazom na črty:

- na akých základoch je postavený prieskum,
- ako ľahko sa používajú a navrhujú,
- aké majú funkcie.

Tabuľka 1
Porovnanie vybraných web aplikácií

	Zoomerang	**MySurveyLab**	**FluidSurveys**	**SurveyMonkey**
Neobmedzené prieskumy	Áno	Áno	Áno	Áno
Šablóny	Áno	Nie	Nie	Áno
Vetvenie alebo preskočenie	Áno	Áno	Áno	Áno
Jazyková podpora	Áno	Nie	Nie	Áno
Zatvorenie stránky	Áno	Áno	Áno	Áno
Duplicitný prieskum	Áno	Áno	Áno	Áno
Prispôsobenie	Áno	Nie	Nie	Áno
Upload e-mailových kontaktov	Áno	Áno	Nie	Áno
Zverejnenie url	Áno	Áno	Áno	Áno
Publikovanie e-mailom	Áno	Áno	Nie	Áno
Export do formátu CSV	Áno	Áno	Nie	Áno
Voľný plán	Áno	Áno	Áno	Áno
Spustenie plánu (mesačne)	19 dolárov	9 dolárov	19 dolárov	35 dolárov

D. Stručný popis možností produktu SurveyMonkey

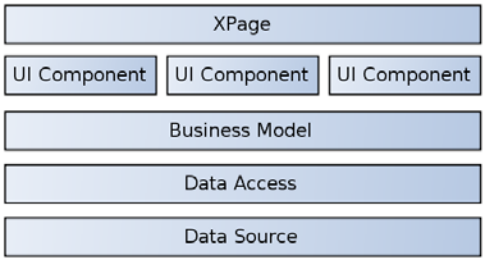
Prieskum SurveyMonkey [2] umožňuje rýchlo vytvoriť anketu alebo prieskum bez akéhokoľvek špeciálneho softvéru. Pomocou online služby, je možné poskytnúť odkaz na prieskumy v e-maile alebo na *w*ebových stránkach a respondenti môžu predložiť svoje odpovede anonymne. Bezplatný prieskum je možný, ale je obmedzený na 10 otázok a 100 odpovedí za mesiac.

Výhody prieskumu SurveyMonkey:

- **Bez obmedzenia** - prieskumy s neobmedzeným počtom otázok,
- **Preskočené/podmienené logiky** - nastavenie automatického vynechanie otázok, ktoré nie sú použiteľné pre respondentov, na základe ich odpovedí na ostatné otázky,
- **Náhodné výbery odpovede** - usporiadanie volieb v rámci otázky môže predstavovať neúmyselné skreslenia odpovede,
- **Filtrovanie výsledkov** - nájdenie vzorov typizovaných odpovedí vo výsledkoch,
- **Zdieľanie výsledkov** - zobrazenie ďalších aj selektívnych výsledkov bez prístupu iným k účtu,
- **Sťahovanie výsledkov** - všetky dáta možno stiahnuť do počítača pre ďalšiu analýzu, napr. do Excelu,
- **Zaručenie súkromia -** otázky a odpovede zostanú súkromné.

II. NÁVRH A IMPLEMENTÁCIA INFORMAČNÉHO SYSTÉMU

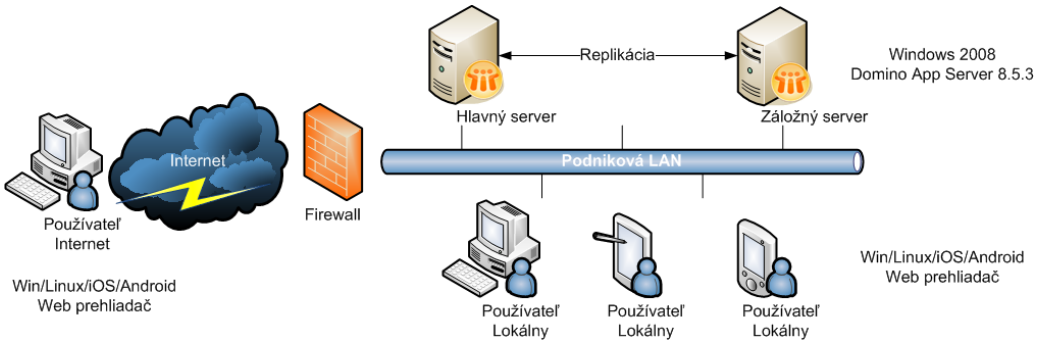
Článok prezentuje časti informačného systému, jednoduchý softvér, ktorý umožnuje vytvárať online prieskumy pre správcu systému a vypĺňať sprístupnené prieskumy pre desiatky online používateľov [4]. Je dostupný na internete v testovacej verzii a realizovaný v prostredí IBM Lotus Domino R8.5 ako klient – server aplikácia Nasledujúce obrázky predstavujú softvérovú architektúru Xpage aplikácie (Obrázok 1) a infraštruktúru navrhovaného systému (Obrázok 2).



Obrázok 1 Popis architektúry Xpage aplikácie

Interný, lokálny používateľ komunikuje priamo zo serverom, externý prostredníctvom

internetu. Prvý server je hlavný server, ktorý štandardne funguje s aplikáciou; druhý server je záložný, v prípade výpadku hlavného servera. Medzi servermi prebieha replikácia, pri ktorej sa údaje zapisujú na jeden aj druhý server v niekoľko minútových intervaloch prípadne aj hneď v závislosti od nastavenia.

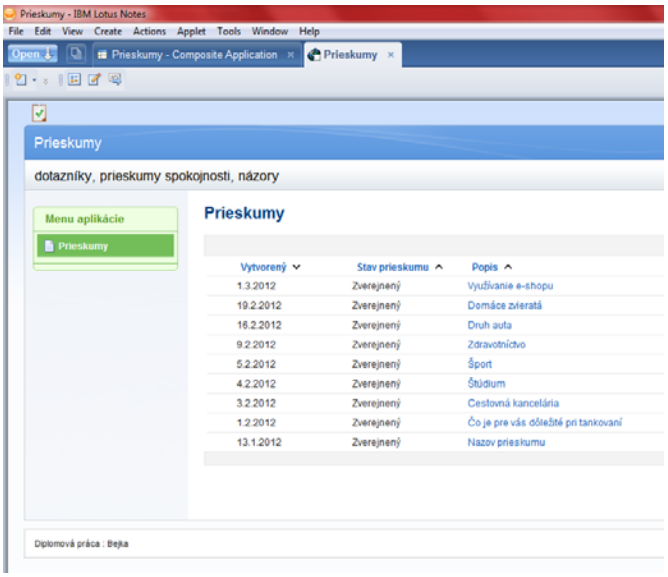


Obrázok 2 Popis infraštruktúry riešenia

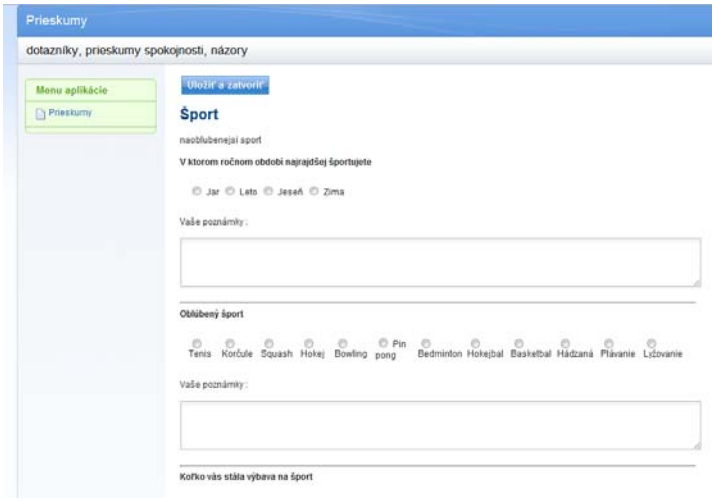
Systém poskytuje nasledovné funkcie:

- Vytvorenie prieskumu - vytvorenie otázok prieskumu pod hlavičkou prieskumu,
- Uloženie otázky i prieskumu,
- Zadanie dátumov publikovateľnosti,
- Zadanie módov dostupnosti – testovacia a ostrá verzia,
- Stav prieskumu – Rozpracovaný/Uzatvorený/Zverejnený,
- Zoznam všetkých aktívnych i neaktívnych prieskumov,
- Súhrn výsledkov prieskumu,
- Export odpovedí – napr. do excelovskej tabuľky,
- Vyplnenie a odoslanie prieskumu používateľom,
- Filtrovanie uskutočnených prieskumov.

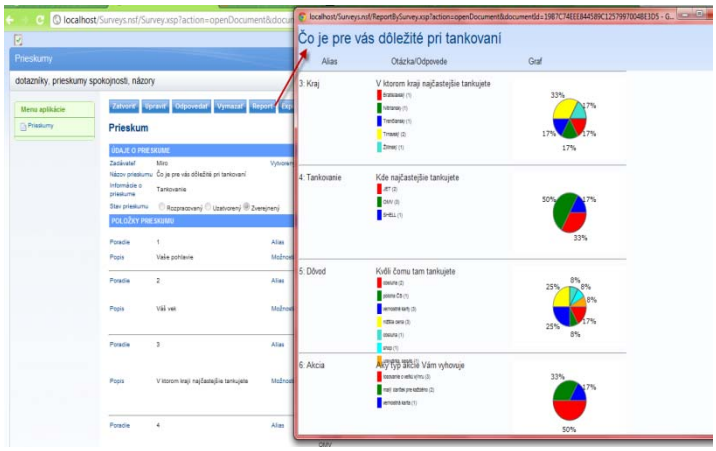
Nasledujúce obrázky obrazoviek implementovaného systému ilustrujú grafické užívateľské rozhranie a základné funkčné vlastnosti systému (pohľad na dostupné prieskumy, detail prieskumu, report odpovedí na prieskum).



Obrázok 3 Pohľad na prieskum



Obrázok 4 Pohľad na detail prieskumu



Obrázok 5 Report odpovedí v danom prieskume

III. ZÁVER

Rozvoj riešení pokrývajúcich online prieskumy v súčasnej dobe má a v najbližších niekoľkých rokoch bude mať ešte väčší význam. Pre zber a získanie spätnej väzby respondentov je tento typ sofvérových riešení v súčasnej dobe už konkurenčnou nevyhnutnosťou. Predpovedaná budúcnosť informačných technológii a nových foriem ich poskytovania je vhodnou oblasťou pre neustále rozširovanie znalostí v tejto oblasti.

POĎAKOVANIE

Tento článok bol podporený projektom KEGA 021TUKE-4/2012 (50%) a projektom Rozvoj centra informačných a komunikačných technológií pre znalostné systémy (číslo projektu: 26220120030) (50%) na základe podpory operačného programu Výskum a vývoj financovaného z Európskeho fondu regionálneho rozvoja a v spolupráci so spoločnosťou PosAm, s. r.o.

LITERATÚRA

[1] *Comparing Survey Applications*. In Listio , dostupné na internete http://www.listio.com/reviews/2009/01/comparing-survey-applications-zoomerang-surveymonkey-fluidsurveys-and-mysurveylab/.

[2] K. Jo'Shea, *Survey Monkey*. In wiki itap purdue, dostupné na internete.https://wiki.itap.purdue.edu/display/INSITE/SurveyMonkey

[3] V. M. Sue, L. A. Ritter, *Conducting online Surveys*. California: Sage Publications,2007. 1-7s. 10-13s. ISBN-13: 978-1-4129-3753-5.

[4] B. Petríková, SFA - IS pre online prieskumy spokojnosti, Diplomová práca, 2012, FEI TUKE, Košice.

Použitie prostriedkov umelej inteligencie pri riešení stability pohybu robota Nao

Marián POLEK, Ján VAŠČÁK

Centrum informačných a komunikačných technológií, Fakulta elektrotechniky a informatiky, Katedra kybernetiky a umelej inteligencie, Technická univerzita v Košiciach, Slovenská republika

marian.polek@student.tuke.sk, jan.vascak@tuke.sk

***Abstrakt* — Hlavnou témou diplomovej práce je použitie prostriedkov umelej inteligencie pri riešení stability pohybu robota Nao. Práca je rozdelená na 4 kapitoly, obsahuje 30 obrázkov a 12 tabuliek. V teoretickej časti práce sú opísané niektoré známe metódy použitia stability pri humanoidných robotoch a v praktickej časti je zdokumentovaný návrh a implementácia riešenia vlastnej metódy stability robota Nao. Jadro práce tvoria 2. a 3. kapitola, kde sa nachádzajú experimenty z implementovanej a pôvodnej b-human metódy stability a ich porovnanie, čo vlastne bolo aj cieľom tejto diplomovej práce.**

***Kľúčové slová* — stabilita, humanoidný robot, Nao, pohyb, chôdza, kopnutie**

I. Úvod

V dnešnej dobe rozlišujeme tri kategórie chodiacich dvojnohých robotov tzv. humanoidných robotov. Do prvej kategórie patria roboti, ktorých rýchlosť chôdze je príliš malá na to, aby sa robot dostal do stavu, v ktorom by stratil stabilitu pri chôdzi, do stavu nerovnovážnosti. Takýto systém sa dá úplne popísať. Roboti, ktorí majú dynamický pohyb a o niekoľko stupňov voľnosti viac ako roboti v prvej kategórii, patria do druhej kategórie za predpokladu, že majú dostatočne veľké chodidlo. Ich Zero moment point (ZMP) je v podporovanom polygóne. V poslednej kategórii sú roboti, ktorí majú len v nohách minimálne 3 - 4 stupne voľnosti, a tým pádom sú aj najzložitejší v otázke riadenia a udržania stability. V prípade, že je v takomto robotovi použitá komplexná metóda udržania stability, je robot schopný dosiahnuť aj pomerne vysokú rýchlosť chôdze. [1]

Stabilita humanoidných robotov záleží od veľa atribútov. Môžeme spomenúť napríklad rýchlosť chôdze, počet stupňov voľnosti ako v nohách, tak isto aj v celom tele alebo typ zeme, na ktorej sa robot pohybuje. V prípade vytvorenia algoritmu na stabilitu, ktorý dokáže využiť aj pasívne časti robotov ako napr. senzory alebo kamera, môžeme dosiahnuť ešte lepšie výsledky. Samozrejmosťou je ale úplné poznanie robota, s ktorým pracujeme.

Keď je dvojnohý robot v pohybe (chôdza), pôsobia naň zotrvačné sily vyvolané gravitáciou a zrýchleniami a spomaleniami zotrvačných síl (inertial force) pri samotnej chôdzi. Výslednica týchto síl sa nazýva celková zotrvačná sila (total inertial force) a má pôsobisko v ťažisku robota. Kontakt robota s podložkou vyvoláva reakčnú silu (floor reaction). V prípade, že robot udržiava perfektnú rovnováhu, sú vektory celkovej zotrvačnej sily a reakcie od podložky v jednej osi. Ak sa tieto sily dostanú mimo svojho osového zarovnania (nerovnosť povrchu), dochádza k strate rovnováhy a generuje sa padacia sila (falling force) [3]

Chôdza dvojnohých robotov alebo mechanizmov je už dlhé roky pozornosťou výskumníkov a profesorov. Teoretické štúdie a projekty sú odsimulované na rôznych typoch robotov od najjednoduchších až po tie najzložitejšie. Bez odhľadu na to, aké sú zložité a koľko stupňov voľnosti majú, každý takýto systém sa skladá z niekoľkých charakteristík: a) schopnosť otočenia na jednej nohe b) vzorkovanie chôdze, opakovanie rovnakých krokov symetrických c) správna zmena dvoch fáz, keď robot stojí na jednej nohe a keď robot stojí na dvoch nohách. Všetky kĺby robota sú ovládané motormi. Spojenie chodidla a zeme sa takisto dá považovať za kĺb, ktorý ale nie je ničím poháňaný. Tento kĺb je veľmi dôležitý pre stabilitu celého systému. Noha nemôže byť ovládaná priamo, a preto sa musí ovládať telo a ruky. Preto celkový indikátor správania sa mechanizmu je vlastne jeden bod, kde všetky sily pôsobiace na mechanizmus sú

nahradené jednou silou. Tento bod je definovaný ako Zero Moment Point (ZMP). Typy chôdze sa preto navrhujú v závislosti na tento bod.

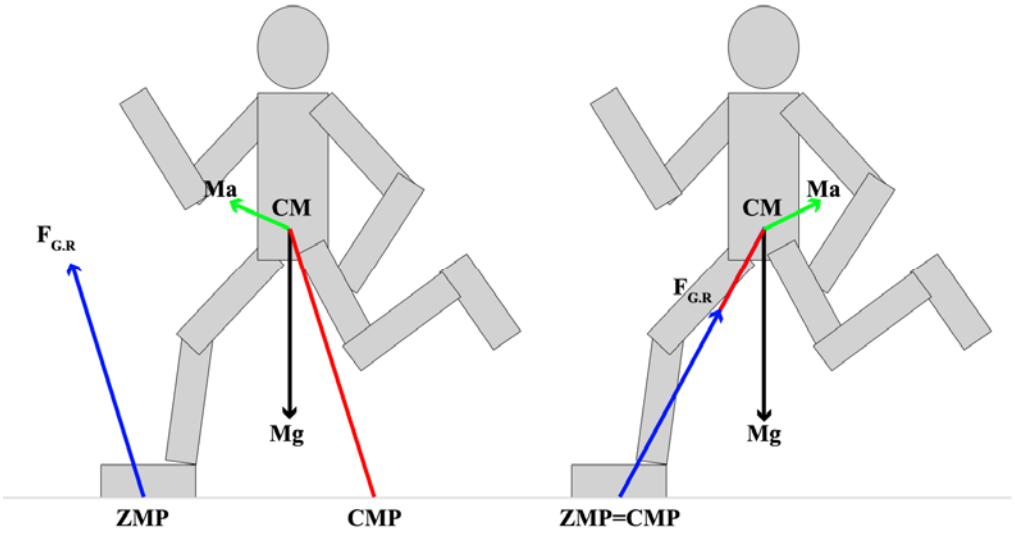
II. REFERENČNÉ BODY NA ZEMI

Existuje niekoľko pozemných referenčných bodov použitých na pohybovú identifikáciu a kontrolu v biomechanike a nožnej robotike. Umiestnenie týchto referenčných bodov vo vzťahu k sebe a vzhľadom k oblasti pozemnej podpory poskytuje dôležité miestne a niekedy aj globálne charakteristiky celotelového pohybu, ktoré slúžia ako kritériá pre buď fyzické alebo požadované pohybové vzory.

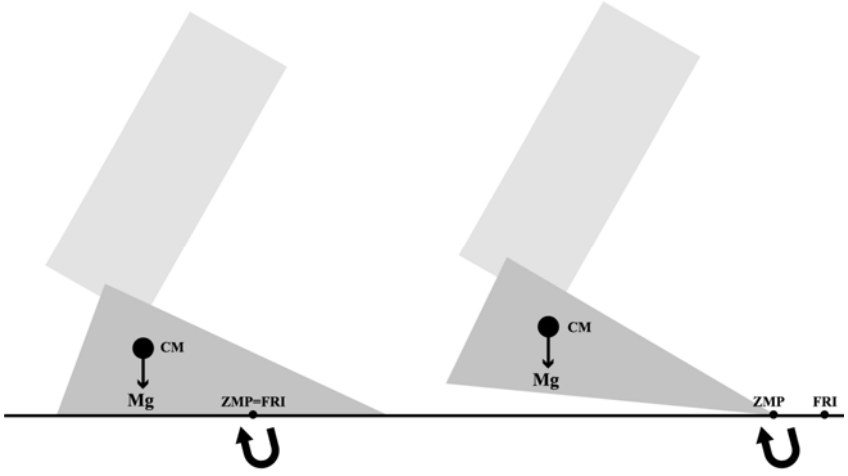
ZMP (Zero moment point) je bod, v ktorom súčet všetkých momentov síl pôsobiacich na systém je rovný nule.

CMP (Centroidal moment pivot) je bod, kde priamka rovnobežná so zemskými reakčnými silami prechádzajúca ťažiskom robota (CM) sa dotkne zemského povrchu. Na **Obr. 1** , vpravo, sa CMP zhoduje so ZMP, keďže reakčné sily zeme smerujú cez bod CM.

FRI (Foot rotation indicator) je bod, kde zemské reakčné sily (sily vyvýjané zemou na telo robota v kontakte s ním) tlmia silu chodidla proti akcelerácii. Ak je noha pevne na zemi, FRI bod je zhodný s bodom ZMP. **Obr. 2** zobrazuje bod v dvoch rozmeroch, ale FRI bod môže byť aplikovaný aj v troch rozmeroch hľadania, vrátane informácií o otáčaní chodidla súčasne v sagitálnej a koronálnej rovine.



Obr. 1 Porovnanie ZMP a CMP



Obr. 2 Porovnanie ZMP a FRI

III. POJEM ZMP

Interpretácia č.1.: ZMP je definované ako bod na zemi, v ktorom sieť momentov inertných síl a gravitačných síl nemá žiadnu zložku, silu pozdĺž horizontálnych osí.

Interpretácia č.2.: P je bod, kde $T_x = 0$ a $T_y = 0$, T_x, T_y reprezentujú momenty okolo x-ovej a y-ovej osi generované reakčnou silou Fr a reakčným krútiacim momentom Tr. Bod P je definovaný ako ZMP. Ak tento bod je v podporovanom polygóne, systém je stabilný.

Častokrát je ZMP zle definovaný, napríklad že ZMP mimo podporovaného polygónu nemá zmysel, ale pravda je, že ZMP mimo podporovného polygónu neexistuje, pretože hneď ako sa dostane mimo podporovaného polygónu, robot spadne na zem. Taktiež je veľmi dôležité ujasniť si pojmy center of pressure (CoP) ťažisko a ZMP. Tlak medzi chodidlom a zemou sa dá vždy nahradiť silou pôsobiacou na ťažisko. Ak táto sila balancuje, vyrovnáva všetky aktívne sily pôsobiace na mechanizmus počas pohybu (inertné, gravitačné, momenty a centrifúgy), potom je to ZMP. Ak je teda krok dynamicky rovnovážny, ťažisko a ZMP sú v jednom bode. Ak krok nie je dynamicky rovnovážny, ZMP neexistuje a mechanizmus kolabuje.

IV. PERSPEKTÍVNE ÚLOHY

Očakávania od robotov stále rastú. Už dnes majú ľudia ambície použiť v robotoch komunikáciu s človekom za účelom zabrániť kritickým situáciám. Všimnime si niektoré vylepšenia do budúcnosti:

- noha a chodidlo, ktoré sa viac približuje realistickej nohe kvôli zvýšeniu stability
- lepšia pružnosť v kĺboch, hlavne kolenných v prípade dopadu nohy na zem, čo je v niektorých materiáloch považované za doménu humanoidnej robotiky
- plynulejší prechod medzi jednotlivými fázami robota ako sú chôdza, chôdza po schodoch, vyhýbanie sa prekážkam alebo zrýchlenie chôdze.
- špeciálna výzva je chôdza nezávislá na ruky, napríklad zobrať pohár zo stola pri chôdzi bez zastavenia
- rovnováha robota nielen na statických podlahách, ale aj na mobilných a dynamických systémoch s rôznymi stupňami voľnosti.

V doteraz pomenovaných systémoch existuje simulátor GHDS (General Human/Humanoid Dynamic Simulator). Jeho najväčším cieľom je dosiahnuť rovnaké výsledky v reálnom svete a simulátore. To znamená, že roboti budú musieť dosiahnuť také výsledky, aby boli pripravení aj na vysoké požiadavky reálnych aplikácií. [2]

Výskumom špeciálnych typov humanoidných robotov sa zaoberajú napríklad aj tímy a výskumníci na celom svete, ktorí neustále vyvíjajú robota Nao, s ktorým sa prezentujú na celosvetových súťažiach v robotickom futbale v Standart Platform League Robo Cup. Jedným z týchto tímov je B-human, ktorý je v súčasnosti najúspešnejším tímom v robotickom futbale.

V. B-HUMAN TYPY KOPOV

B-human je názov pre nemecký tím vysokoškolských študentov, ale aj profesorov z Univerzity Bremen, fakulty matematiky a informatiky. Tento tím je výnimočný tým, že už niekoľko krát zvíťazil na majstrovstvách v robotickom futbale. Robotický futbal je veľmi podobný klasickému futbalu čo sa týka pravidiel, no je veľkosťou prispôsobený robotom Nao od firmy Aldebaran robotics. K tak veľkému úspechu ako napríklad víťazstvo 11:1 v jednom z ich zápasov prispieva hlavne veľmi dobre naprogramovaný celkový systém. Rozhodne k tomu prispieva aj ich systém kopu, stabilitu nevynímajúc.

Pri kope je dôležité udržať rovnováhu, nespadnúť. Každý kop je definovaný ako séria statických uhlov kĺbov. Každá séria je kľúčová. Všetky kĺby medzi dvoma kľúčovými sériami sú interpolované. Hlavná nevýhoda takéhoto prístupu je v jeho flexibilite. Keď raz pohyb začne, nedá sa ovplyvniť ani zastaviť, prijať nové informácie. Podobné prístupy sú vo viacerých robotických tímoch. B-human obohatil tento prístup o možnosť balancovania na kompenzovanie vonkajších otrasov. Kľúčové série uhlov kĺbov sú definované ako karteziánske pozície končatín, ktoré sú interpolované. Skutočné uhly kĺbov sú vypočítané pomocou inverznej kinematiky. Systém obsahuje modul na stabilizáciu, ktorý je založený na ťažisku a senzoroch spätnej väzby uzavretého PID regulátora.

Pohybový modul je vyvinutý ako modul, ktorý môže byť použitý paralelne s ostatnými modulmi ako napríklad s modulom kráčania. Hlavná myšlienka je urobiť jednoduché krivky pohybov zložitejšími. Celý pohyb je potom rozdelený do takzvaných fáz, ako napríklad zdvihnutie nohy alebo kopnutie správnou nohou. Každá fáza má svoju dobu trvania a každá končatina obsahuje svoju krivku pohybu. Taktiež je nutné zabezpečiť, aby prechod medzi fázami bol plynulý a hladký. Modul pohybu prijme zoznam pozícií všetkých kĺbov, ale aj

napríklad informáciu o polohe lopty ako vstup. Ako výstup je samotný pohyb v závislosti na okolnosti.

Základná štruktúra modulu pohybu je zložená z niekoľkých krokov. Najskôr zistí ID pohybu, ktorý má byť vykonaný. Potom, na začiatku každej fázy sa aplikujú dynamické body na aktuálnu fázu a inicializujú sa trajektórie. Rovnako je zabezpečená aj plynulosť pohybu. Následne sa získajú pozície končatín v aktuálnom čase a aplikuje sa inverzná kinematika na výpočet uhlov kĺbov pre pohyb ostatných končatín na dosiahnutie stability, rovnováhy. [4]

Princíp stability robota Nao v robotickom futbale je čiastočne porovnateľný ako je riešený v tejto diplomovej práci, aj keď v nej navrhované metódy fungujú na odlišnom princípe.

VI. NAVRHNUTÁ METÓDA

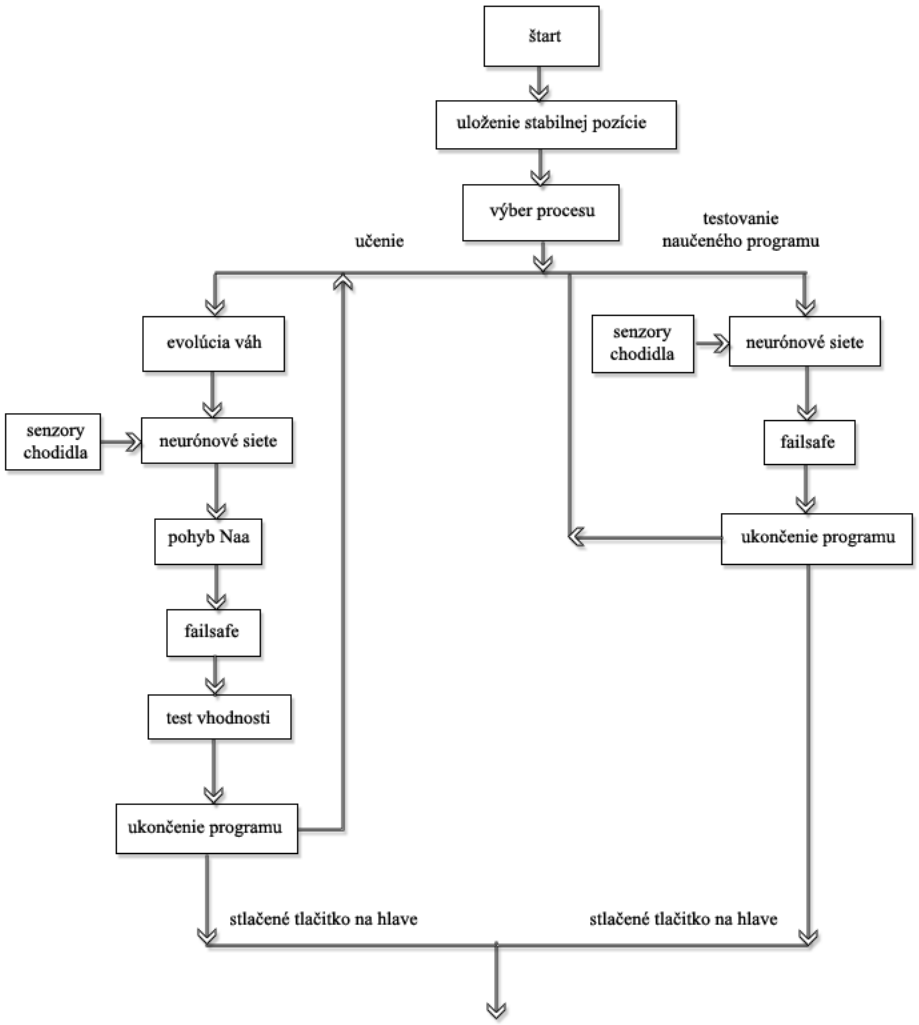
Udržanie rovnováhy humanoidného robota je pomerne náročný proces, ktorý sa samozrejme dá riešiť rôznymi spôsobmi. Implementovaný spôsob udržania rovnováhy humanoidného robota na jednej nohe je zložený z dvoch veľkých celkov a je pomerne jednoduchší ako stabilita riešená v tíme B-human. Riešenie stability robota Nao je navrhnuté za pomoci tlakových senzorov na spodku chodidla robota Nao a následným vykonaním pohybu stability za pomoci neurónových sietí.

Prvý celok je samotné vykonanie pohybu jednotlivých častí tela tak, aby robot Nao nespadol na zem, aby nestratil stabilitu, rovnováhu. To, aby udržal rovnováhu, môžu zabezpečiť napríklad ruky, noha, ktorá nedrží telo, ale je voľná vo vzduchu, samotný trup ale aj hlava. Program, ktorý bol navrhnutý, použil na korigovanie rovnováhy voľnú nohu a trup. Týmito dvoma časťami robot hýbe, aby udržal rovnováhu.

Druhý bod samotného programu je učenie. Učenie je veľmi dôležité z pohľadu správneho pohybu vyššie spomínaných častí tela na korigovanie rovnováhy. To znamená učenie robota Nao naučí, kedy, kde a o koľko sa majú jednotlivé časti tela pohnúť.

Tieto dva najväčšie celky programu sú úzko prepojené. Pri spustení programu robot Nao nevykonáva nič okrem toho, že sa postaví na jednu nohu. Najskôr sa v členku nakloní doľava a potom zdvihne nohu. Nie je možné urobiť tento postup naopak, lebo pri zdvihu nohy robot spadne doprava. Preto je potrebné najskôr v členku celého robota Nao nakloniť doľava, aby sa ťažisko robota premiestnilo trošku doľava a až potom následne zdvihnúť nohu. Takýto postup, algoritmus, môže byť pre estetickosť doplnený o nastavenie rúk do polohy lastovičky. To však nie je potrebné.

Po spustení a pohybe Naa do pozície postoja na jednej nohe robot čaká na používateľa. Tromi dotykovými tlačidlami na hlave sa vykonajú nami požadované operácie. Jedným tlačidlom sa vykoná operácia uloženia stabilnej pozície, to znamená uloženia senzorov z chodidla, ktoré sú v stabilnej polohe a druhým tlačidlom sa začne proces učenia. Tretie nám slúži na to, keď práve nechceme učiť Naa, ale iba vykonať stabilizáciu, pustiť Naa na ostro. Na **Obr. 3** je zobrazený hrubý návrh algoritmu učenia. Evolučný algoritmus vytvára váhy, ktoré sú použité v neurónových sieťach. Neurónové siete vypočítavajú aký pohyb má Nao urobiť.



Obr. 3 Algoritmus navrhnutej metódy

VII. IMPLEMENTOVANÝ ALGORITMUS

Ako implementačný algoritmus je použitý navrhnutý algoritmus. Kvôli B-human softwaru musel byť ale trochu pozmenený. Prvá zmena spočívala v tom, že robot pri stabilizácii nepoužíval nohu, ktorá je vo vzduchu a trup, ale obe horné končatiny. Konkrétne štyri časti horných končatín LShoulderPitch, RShoulderPitch, LShoulderRoll a RShoulderRoll. Namiesto dvoch boli použité štyri neurónové siete. Nohou, ktorú mal Nao vo vzduchu, nemohol zabezpečovať rovnováhu, kedže ňou kope do lopty. Trup priamo súvisí s oboma nohami a nie je možné hýbať trupom pri kope do lopty.

Ďalšou úpravou bolo, že proces učenia vznikol na navrhnutom systéme a do implementovaného algoritmu sa vložila časť už naučeného systému.
Nahradená bola funkcia simpleCalcArmJoints() novou metódou. Funkcia simpleCalcArmJoints() sa nachádza vo funkcii calcJoints() a stabilitu Naa rieši pomocou inverznej kinematiky. Táto funkcia je pri B-human volaná dvakrát pre ľavú a dvakrát pre pravú ruku.

Algoritmus implementovanej metódy pracuje najskôr tak, že zistí hodnoty na všetkých štyroch tlakových senzoroch chodidla nohy, ktorá stojí pevne na zemi. Tieto hodnoty sú zapísané do pomocných premenných. Pomocou štyroch neurónových sietí sa vypočítajú hodnoty, ktoré sa následne použijú na pohyb horných končatín. Pre správne vloženie hodnôt do neurónových sietí bolo potrebné tieto zistené hodnoty najskôr predeliť číslom 4000 z dôvodu použitia sigmoidálnej aktivačnej funkcie, pretože hodnoty z tlakových senzorov dosahujú rozsah od 0 po 4. Pre silnejší pohyb boli potom výsledné hodnoty vynásobené číslom 100. Konkrétne riešenie pre pohyb horných končatín spočíva v zistení aktuálnej pozície jednotlivého

kĺbu a pripočítanie výslednej hodnoty vynásobenej číslom 100. Čísla 4000 a 100 po niekoľkých opakovaných pokusoch vyšli ako najúčinnejšie.

VIII. VÝSLEDKY EXPERIMENTOV

Výsledky implementovanej metódy sú zaznamenané v tabuľkách. Sila kopu je udaná v milisekundách, keďže sa jedná o trvanie švihu nohy.

Vzdialenosť je dĺžka od robota k dopadu vykopnutej lopty. Je udávaná v centimetroch.

Ďalšie veličiny sa týkajú hodnôt nameraných uhlov horných končatín. Celkový súčet týchto hodnôt udáva pozíciu rúk k telu. Čím je suma týchto hodnôt menšia, tým sú horné končatiny robota viac pri tele a tým pádom je robot ustálený a pripravený na ďalšiu fázu pri robo-futbale. Ak je sila kopnutia príliš veľká, dôjde k jemnému otočeniu robota. Táto hodnota je udávaná v stupňoch. Čím menšie je natočenie, tým je robot stabilnejší a jeho poloha sa pri futbale nezmení. V prípade veľkej sily kopnutia by mohlo dôjsť aj k strate rovnováhy a robot by mohol spadnúť. Pri meraní so silou 200 a 100 ms neboli namerané konkrétne hodnoty, nakoľko veľkosť ihriska je v porovnaní s reálnym futbalovým ihriskom oveľa menšia. Preto boli hodnoty vyššie ako 250 cm zaokrúhlené na hodnotu 250 cm.

Tabuľka 1
Výsledky imeplementovanej metódy

Implementovaná metóda					**Priemer**
Sila (dĺžka v ms)	500	400	300	200	
Vzdialenosť (cm)	69,8	111,6	162,8	250	**148,55**
Súčet uhlov kĺbov	409,56	433,94	446,62	462,98	**438,275**
Natočenie (stupne)	0	0	3,6	9,2	**3,2**

Pri metóde B-human bol zvolený rovnaký postup experimentov ako aj zobrazenia výsledkov.

Tabuľka 2
Výsledky B-human metódy

B-human metóda					**Priemer**
Sila (dĺžka v ms)	500	400	300	200	
Vzdialenosť (cm)	49,6	82,4	178,2	244,4	**138,65**
Súčet uhlov kĺbov	611,9	612,72	610,74	611,3	**611,665**
Natočenie (stupne)	0	0	3	9,2	**3,05**

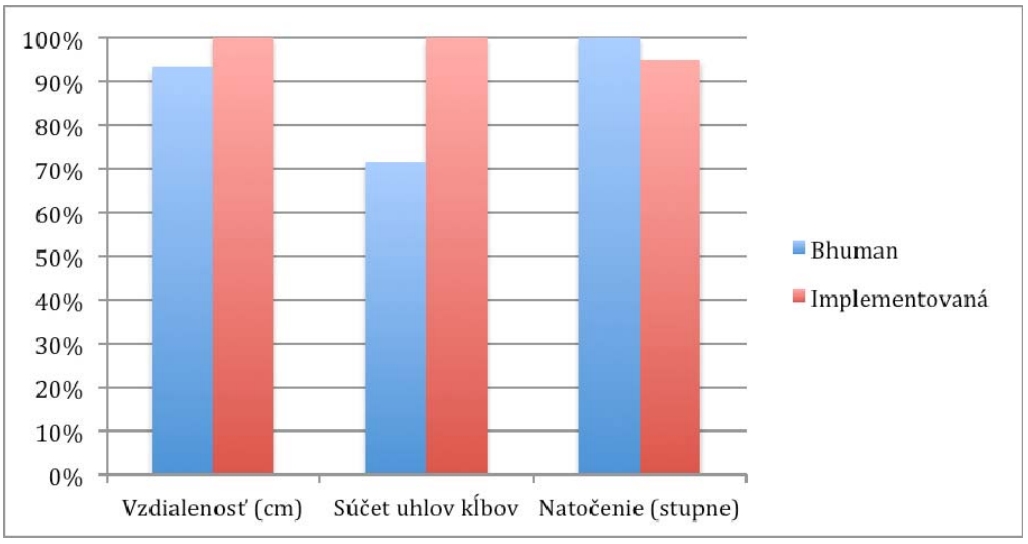
IX. POROVNANIE VÝSLEDKOV OBOCH METÓD

V grafe je vyobrazené percentuálne porovnanie jednotlivých prvkov merania. Grafy sú rozdelené podľa sily kopu. Z výsledkov vyplýva, že pri sile 500 ms je implementovaná metóda lepšia z hľadiska vzdialenosti dokopnutia a pozície horných končatín. V tomto prípade sa robot nenatočil vôbec, z čoho vyplýva, že bol stabilný. Veľmi podobné výsledky sú dosiahnuté aj pri sile 400 ms. Pri sile 300 ms výsledky meraní boli lepšie u B-human, a to vo vzdialenosti dokopnutia a natočenia robota. Napriek lepším výsledkom u B-human pri natočení robota sú však tieto hodnoty pomerne rovnaké, keďže natočenie bolo v priemere trojstupňové. Pri sile 200 ms sú výsledky implementovanej metódy lepšie ako metódy B-human. Merania pri sile 100 ms nie sú zaznamenané, nakoľko boli skoro identické s hodnotami pri sile 200 ms.

Celkovo ale je možné zhodnotiť, že implementovaná metóda stability robota Nao pri kope je lepšia s porovnaním metódy B-human. Napriek tomu ale robot nestratil rovnováhu ani pri jednej z experimentálnych metód. Z toho vyplýva, že obe metódy sú pomerne stabilné.

Tabuľka 3
Porovnanie oboch metód

	B-human metóda	**Implementovaná metóda**	**percentuálny rozdiel**
Vzdialenosť (cm)	138,65	148,55	**6,66**
Súčet uhlov kĺbov	611,665	438,275	**39,56**
Natočenie (stupne)	3,05	3,2	**-4,92**



Obr. 4 Grafické porovnanie výsledkov oboch metód

POĎAKOVANIE

Táto práca bola vytvorená realizáciou projektu Rozvoj Centra informačných a komunikačných technológií pre znalostné systémy (kód ITMS projektu: 26220120030) na základe podpory operačného programu Výskum a vývoj financovaného z Európskeho fondu regionálneho rozvoja.

LITERATÚRA

[1] Hataitep Wongsuwarn, and Djitt Laowattana (2006). Neuro-Fuzzy Algorithm for a Biped Robotic System. In: PROCEEDINGS OF WORLD ACADEMY OF SCIENCE, ENGINEERING AND TECHNOLOGY VOLUME 15 OCTOBER 2006

[2] M. Vukobratovic, B. Borovac, "ZERO-MOMENT POINT — THIRTY FIVE YEARS OF ITS LIFE," October 2003

[3] Sciranka, M.: Lokomočné mechanizmy kráčajúcich robotov, In: 1st Annual International Ph.D. Conference AVS/PMA, s.: 134 – 138, Praha 2007McMahon T., Bonner J.: On size and life, New York 1983

[4] Deutsches Forschungszentrum für Künstliche Intelligenz, B-Human Team Report and Code Release 2011, November 3, 2011

Optimalizácia výrobného procesu prostredníctvom štíhlej výroby

Lukáš POPIK, František BABIČ

Katedra kybernetiky a umelej inteligencie, Fakulta elektrotechniky a informatiky,
Technická univerzita v Košiciach, Slovenská republika

lukas.popik@student.tuke.sk, frantisek.babic@tuke.sk

***Abstrakt* — Tento článok približuje problematiku štíhlej výroby a jej implementáciu do reálnej firmy s cieľom optimalizovať vybraný výrobný proces. Jednou z mnohých metód štíhlej výroby je mapovanie tokov hodnôt, ktoré predstavuje nástroj na odhalenie nedostatkov v procese a návrh riešení vedúcich k zvýšeniu efektivity výroby. Na tento účel je možné využiť špecializovaný softvér, v tomto prípade SigmaFlow. Úspešnosť navrhnutého riešenia bude vyhodnotené z pohľadu zvýšenia efektivity skúmaného výrobného procesu.**

***Kľúčové slová* — Optimalizácia, Mapovanie tokov hodnôt, Štíhla výroba**

I. Úvod

Účinným spôsobom ako v zložitej ekonomickej dobe minimalizovať náklady je identifikovať miesta, v ktorých dochádza k ich plytvaniu a odstrániť ich. Na tento účel je možné použiť proces mapovania tokov hodnôt prostredníctvom vytvorenej mapy nielen pre materiálový ale aj informačný tok vo výrobnom procese. Prvým krokom bude vytvorenie mapy súčasného stavu, ktorá zachytáva aktuálny stav výrobného procesu, a na jej základe budú identifikované miesta pre zefektívnenie skúmaného procesu. Tieto zmeny budú dokumentované mapou budúceho stavu, ktorej rozdiel v efektivite oproti mape súčasného stavu vyjadruje požadované zlepšenia. V tomto prípade bola stanovená hranica zlepšenia minimálne o 10%.

Tento článok je rozdelený do piatich hlavných častí, kde v prvých dvoch je stručne predstavená motivácia a história metodiky štíhlej výroby. Tretia sekcia popisuje postup vytvárania máp súčasného a budúceho stavu, na ktorú nadväzuje štvrtá sekcia s prezentáciou dosiahnutých výsledkov. Záver sumarizuje použitý prístup a prínosy navrhnutého riešenia.

II. Koncept štíhlej výroby

A. História štíhlej výroby

Metodika štíhlej výroby má svoje počiatky v päťdesiatych až šesťdesiatych rokoch dvadsiateho storočia vo firme Toyota. Vznikla ako alternatíva k hromadnej výrobe v prostredí, ktoré vyžadovalo vysokú úroveň flexibility a kde chýbali financie na nákladné investície. Otcom výrobného systému Toyoty je manažér Taiichi Ohno (1912-1990), ktorý bol v roku 1947 vedúcim jednej výrobnej jednotky v Toyote. (bežne dostupné zdroje)

Výrobný systém Toyota, TPS, je tvorený dvoma základnými piliermi: JIT, ktorý sa interpretuje ako výroba/dodávka práve včas a Jidoka, teda automatizácia s ľudskou inteligenciou. Pilier JIT (prevzatý z amerických automobilových závodov) znamená, že sa potrebné súčiastky v procese toku dostanú na montážnu linku v pravom čase a v potrebnom množstve. Pilier Jidoka znamená, že stroj simuluje ľudskú vlastnosť, kedy dokáže rozlíšiť zlý produkt - nepodarok od dobrého a v prípade komplikácií a problémov sa výroba automaticky zastaví, prípadne iným vhodným spôsobom sa zamedzí vzniku nepodarku. V päťdesiatych a šesťdesiatych rokoch bola práca Taiichiho Ohna doplnená o výsledky Shigea Shinga (1909-1990) v oblasti znižovania časov na nastavovanie SMED, čo umožnilo vyrábať v menších dávkach. Takto vytvorená flexibilita umožnila Toyote a niektorým ďalším japonským automobilkám vyrábať a dosahovať zisk aj počas ropnej krízy. (bežne dostupné zdroje)

ISBN 978-80-553-0890-6 © 2012 FEI TUKE

III. MAPOVANIE TOKOV HODNÔT

Mapovanie toku hodnôt je jedným z nástrojov štíhlej výroby, ktorý slúži na odhalenie a vizualizáciu procesov, ktoré pridávajú ale aj nepridávajú hodnotu vo výrobných, servisných ale aj administratívnych štruktúrach. Na to, aby bolo možné odhaliť hodnotu pridávajúce a hodnotu nepridávajúce činnosti, je najvhodnejšia vizualizácia formou mapy. Príkladom hodnoty pridávajúcej činnosti je montáž súčiastok a hodnotu nepridávajúcej státie materiálu na sklade alebo transporty. Hodnoty pozdĺž výrobného procesu sa spájajú a vytvárajú toky hodnôt. Je možné rozlíšiť dva druhy tokov hodnôt vo výrobnom procese a to toky materiálové a toky informačné.

A. Mapovanie súčasného stavu

Úlohou mapy súčasného stavu je zdokumentovať to „čo práve vidíme“. Mapa súčasného stavu bola vytvorené podľa nasledujúcich krokov:

1. Výpočet času taktu (takt time).
2. Kontrola procesu od začiatku po koniec.
3. Vyznačenie požiadaviek zákazníka.
4. Mapovanie od konca procesu.
5. Dôraz na tok materiálu.
6. Vyznačenie informačného toku a časovej osi.

Po vykonaní mapovania súčasného stavu bola pomocou príslušných rovníc vyjadrená efektivita výrobného procesu:

$$E_C\,[\%]=(228{,}866/521{,}602)*100=43{,}878\ \%$$

B. Mapovanie budúceho stavu

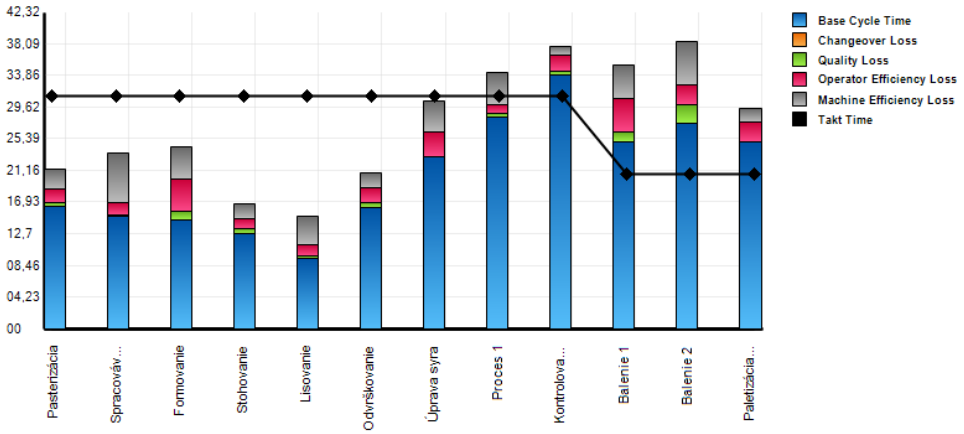
Mapa budúceho stavu predstavuje mapu súčasného stavu s vyznačenými miestami, kde existuje predpoklad pre dosiahnutie zlepšenia z pohľadu času vykonávania alebo spotrebovaných nákladov. Proces tvorby mapy budúceho stavu pozostával z nasledujúcich krokov:

1. Vytvorenie grafu znázorňujúceho časy cyklov a čas taktu.
2. Na základe grafu cyklov a taktu boli identifikované tzv. úzke miesta.
3. Na základe týchto miest boli navrhnuté zmeny zakreslením do mapy budúceho stavu a vypočítané nové hodnoty jednotlivých ukazovateľov, napr. efektivita výrobného procesu so navrhnutými zmenami.
4. Na koniec bola vyčíslená zmena stavu.

IV. DOSIAHNUTÉ VÝSLEDKY

A. Mapovanie tokov hodnôt

Prvý spôsob prezentácie dosiahnutých výsledkov predstavuje graf vyjadrujúci časy cyklov procesov a čas taktu vzhľadom na dopyt zákazníka (Obr.1).



Obr. 1 Graf vyváženia záťaže

Tento graf pozostáva z dvoch osí, kde na horizontálnej sú znázornené názvy vybraných

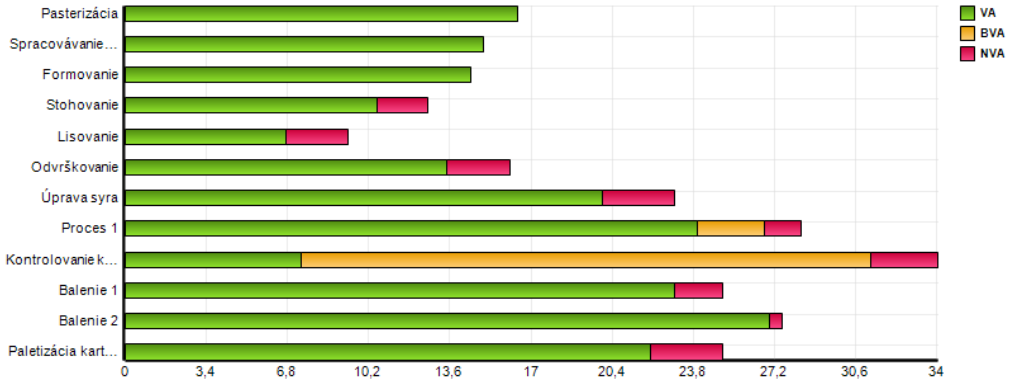
procesov výrobnej linky, ktoré boli predmetom porovnania, a na osi vertikálnej sú číselné údaje, ktoré vyjadrujú dĺžku trvania jednotlivých procesov v minútach. Ďalej z grafu vieme vyčítať čas taktu, ktorý má hodnotu 31,082 minúty a v grafe je znázornený krivkou čiernej farby.

Čas taktu je reprezentovaná krivkou a nie priamkou z dôvodu, že nie každý výrobný proces sa vykonáva v rovnako-zmennej prevádzke. Konkrétne od procesu *Pasterizácie* až po proces *Kontrolovania* výroba funguje na tri zmeny pričom *Balenie 1, Balenie 2* a *Paletizácia* v dvojzmennej prevádzke. Každý proces sa skladá z minimálne troch dielov a maximálne štyroch, ktoré sú farebne rozlíšené a označené v legende grafu. Modrý diel každého procesu vyjadruje čas trvania konkrétneho procesu na jeden kus a označuje sa ako *Base Cycle Time*. Červený diel každého procesu predstavuje stratu efektivity operátora – *Operator Efficiency Loss* a sivý diel stratu efektivity stroja – *Machine Efficiency Loss*. Diel zelený, vyskytujúci sa iba v niektorých procesov, predstavuje stratu na kvalite – *Quality Loss*. Ako vidíme z grafu, oranžový diel – *Changeover Loss* nie je znázornený ani v jednom z našich procesov, a to z dôvodu, že výstupom z výrobného procesu nie sú čiastkové výstupy.

Na základe Obr.1 môžeme povedať, že v trojzmennej prevádzke je úzkym miestom proces *Kontrolovania*, ktorý presahuje čas taktu o 2,918 minúty. V prevádzke dvojzmennej sú úzkymi miestami všetky tri procesy: *Balenie 1, Balenie 2* a *Paletizácia*, pričom *Balenie 1* a *Paletizácia* presahujú čas taktu o 4,279 minúty a *Balenie 2* až o 6,779 minúty.

B. Analýza hodnôt

V oblasti analýzy hodnôt je možné výsledky prezentovať troma typmi grafov. Prvým z grafov analýzy hodnôt podľa veličiny času – *Value Analysis (Activities by Minutes of Time)* je absolútny graf na nasledujúcom obrázku č.2.



Obr. 2 Analýza hodnôt z hľadiska času (absolútny pohľad)

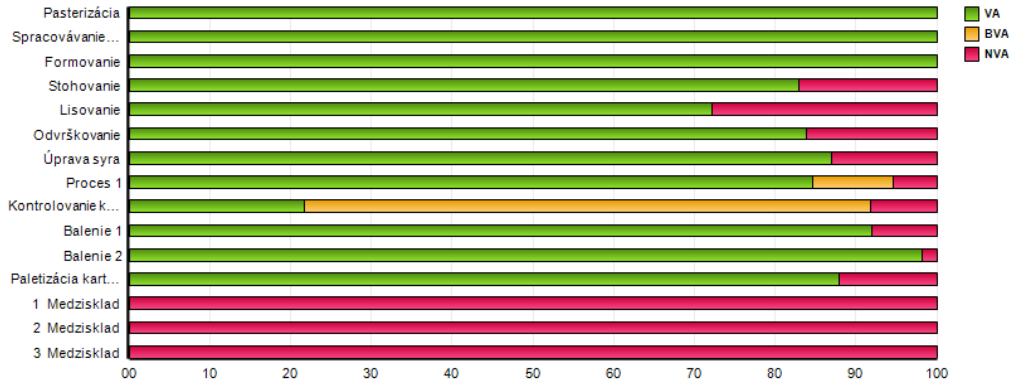
Vertikálnu os tohto grafu tvoria názvy procesov a horizontálnu číselné údaje vyjadrujúce trvanie jednotlivého procesu. Tak ako to popisuje legenda, zelenou farbou je vyznačená tá časť procesu, ktorá má pre zákazníka význam z pohľadu hodnoty odrážajúcej sa vo finálnom produkte – *Value Added*. Farbou červenou je znázornená tá časť procesu, ktorá pre zákazníka význam nemá – *Non Value Added*. Špecifická časť procesu je znázornená farbou oranžovou a predstavuje nehmotnú hodnotu – *Business Value Added.*

V tomto prípade *Business Value Added* tvorí až 70% celkového trvania procesu *Kontrolovania*, čo je možné vidieť na nasledujúcom Obr.3. Tento relatívny pohľad sa od pohľadu absolútneho líši hodnotami na horizontálnej osi, kde sú ako číselné hodnoty uvedené percentá.

Graf reprezentujúci relatívny pohľad predstavuje graf všeobecný. Vyjadruje akú časť daného procesu predstavuje plytvanie – nepridávajúca činnosť. Od absolútneho pohľadu sa nelíši len hodnotami na osi x, ale pribudli aj ďalšie tri procesy. Ide o procesy skladovania, *Medzisklad 1-3*, ktoré sú znázornené len červenou farbou. Znamená to, že tieto procesy sú v celom svojom trvaní hodnotu nepridávajúce, teda predstavujú najväčší zdroj plytvania.

Vzhľadom na súčasné objemy produkcie a fakt, že úzke miesto tvoria procesy *Balenie 1*, *Balenie 2* a *Paletizácia,* jedným z hlavných návrhov na zlepšenie je prechod z dvojzmennej prevádzky na týchto spomínaných troch procesoch na prevádzku trojzmennú. Tým pádom by sa takisto vytvoril priestor na zvýšenie objemu produkcie. V prípade, že by sa tento výrobný proces vykonával v trojzmennej prevádzke, bolo by možné na základe Obr.3 zrušiť *Medzisklad*

3, keďže predstavuje časť hodnotu nepridávajúcu. Dôležitým faktorom pre rozhodovanie je fakt, že niektoré sklady sú špecifickými, keďže v nich prebiehajú jedinečné procesy. Preto by bolo možné zrušiť *Medzisklad 3* len vtedy, ak by nebol skutočne technologicky nevyhnutný. Druhým krokom by bolo nielen zrušenie príslušného transportu pre prechod z trojzmennej prevádzky na dvojzmennú, ale aj automatizácia transportu medzi procesmi *Kontrolovanie* a *Balenie 1* pomocou dopravníka, čo by umožnilo šetriť ľudské zdroje a vytvorila by sa plynulá línia výrobného procesu. Všetky tieto zmeny sú súčasťou mapy budúceho stavu.



Obr. 3 Analýzy hodnôt z hľadiska času (relatívny pohľad)

V. VÝSLEDKY

Cieľom implementácie metodiky štíhlej výroby do prostredia vybranej firmy bola optimalizácia výrobného procesu z pohľadu celkovej efektivity. Toto riešenie je postavené na prístupe mapovania toku hodnôt pomocou vytvorenia mapy súčasného a budúceho stavu v nástroji SigmaFlow. V prípade mapy súčasného stavu bola vypočítaná hodnota efektivity skúmaného výrobného procesu 43,878%. Po zapracovaní navrhnutých zmien a vytvorení mapy budúceho stavu bola vypočítaná nová hodnota efektivity 57,750%. Na základe týchto dvoch hodnôt je možné porovnať obe efektivity, čo predstavuje posledný krok procesu tvorby mapy budúceho stavu:

$$I\,[\%] = |E_C - E_F|$$

$$I\,[\%] = |43{,}878 - 57{,}750| = 13{,}872\ \%$$

Na základe dosiahnutej hodnoty zlepšenia 13,872 % je možné skonštatovať, že stanovený cieľ bol splnený a prostredníctvom navrhnutých zlepšení je možné optimalizovať skúmaný výrobný proces z pohľadu času a nákladov.

POĎAKOVANIE

Tento príspevok vznikol za podpory KEGA projektu č. 065TUKE-4/2011 (50%) a realizáciou projektu Rozvoj Centra informačných a komunikačných technológií pre znalostné systémy (kód ITMS projektu: 26220120030) na základe podpory operačného programu Výskum a vývoj financovaného z Európskeho fondu regionálneho rozvoja (50%).

Návrh a realizácia aplikácie na analýzu výrobných procesov pomocou Petriho sietí

Rudolf REPČÍN, Matej ČOPÍK, Ján JADLOVSKÝ

Katedra kybernetiky a umelej inteligencie, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

rudolf.repcin@student.tuke.sk, matej.copik@tuke.sk, jan.jadlovsky@tuke.sk

***Abstrakt* — Tento článok popisuje návrh a realizáciu aplikácie pre analýzu výrobných procesov modelovaných pomocou Petriho sietí. V úvodnej časti popisuje problematiku výrobných procesov a ich modelovania. V ďalšej časti sú popísané časti a funkcie aplikácie a v poslednej časti sú funkčnosť a možnosti aplikácie demonštrované na príklade modelu výrobnej linky.**

***Kľúčové slová* — výrobná linka, Petriho siete, aplikácia v jazyku C#, modelovanie**

I. ÚVOD

Cieľom tejto práce bolo navrhnúť a realizovať aplikáciu, ktorá by slúžila ako nástroj pre výpočty s maticami Petriho siete, ich editáciu, a analýzu výrobných liniek popísaných Petriho sieťou. Aplikácia môže byť použiteľná pri výpočtoch časov výroby alebo nákladov na výrobu výrobkov, takisto na analýzu stavov Petriho siete a rôzne štatistické výpočty. Mala by slúžiť hlavne ako pomôcka pri optimalizácii výrobných procesov.

II. MODELOVANIE VÝROBNÝCH PROCESOV POMOCOU PETRIHO SIETÍ

Dnešné výrobné systémy sú veľmi komplexné a náklady na zostavenie a údržbu mnohých z nich sú vysoké.

Monitorovanie funkčnej činnosti týchto systémov má významnú úlohu v automatizovanej výrobe, kde na výrobu nedohliada človek osobne. Výber monitorizačných funkcií, ktoré je potrebné identifikovať a monitorovať najmä v zložitých výrobných systémoch je zväčša problémový.

Jedna z efektívnych metód pre výber monitorizačných funkcií je metóda založená na využívaní Petriho sieti (PS). Na opis správania sa systémov možno za najpoužívanejšie považovať PS a ich mnohé modifikácie. Mnohé vlastnosti PS umožňujú konštruovanie hierarchických štruktúr, ktoré sú vhodné pri projektovaní zložitých systémov ako aj modelovanie jednotlivých výstavbových prostriedkov robotizovaných výrobných systémov.

Pre vhodný návrh riadenia pre proces je nutné mať o tomto procese čo najviac informácií. Poznať správanie daného systému v jednotlivých situáciách, poznať jeho požiadavky na riadenie a rovnako aj fyzikálne či iné obmedzenia špecifické pre daný systém. Je dôležité mať jasne stanovené aké sú ciele systému a ktoré akcie majú akú prioritu, aby nedochádzalo v systéme ku kolíziám.

Cieľom požiadaviek na poznanie systému je mať k dispozícii čo najlepší všeobecný model, ktorý čo najdokonalejšie popisuje dynamiku daného systému. Čím je model všeobecnejší, tým lepšie je ho možné analyzovať. Slabou stránkou Petriho sietí je problém komplexity, čiže modely založené na Petriho sieťach majú tendenciu byť príliš veľké aj pri stredne veľkých systémoch.[1]

Modelovanie pomocou Petriho sietí sa uskutočňuje sa na úrovni stavov výrobného procesu. V modeli sú určené činnosti, ktoré sa vykonávajú v systéme, stavy, ktoré predchádzajú týmto činnostiam a stav, do ktorého sa systém dostane po vykonaní činností. Simulovaním tohto stavového modelu pomocou Petriho sietí môžme získať popis správania sa systému.[4]

III. POUŽITÉ TECHNOLÓGIE

Na úpravu zdrojového kódu bolo použité integrované vývojové prostredie Microsoft Visual Studio 2010 od spoločnosti Microsoft. Použitá je platforma.NET Framework 4. Aplikácia je napísaná v jazyku C#.

IV. REALIZÁCIA APLIKÁCIE

Pri tvorbe aplikácie bolo dôležité vychádzať z cieľov, na ktoré ma program slúžiť. Pre jednotlivé ciele bola aplikácia rozčlenená na viaceré časti popísané v nasledujúcich podkapitolách.

A. Editor matíc

Ako vstupy do aplikácie sú potrebné matice ***Pre*** a ***Post*** a vektor počiatočného značenia $\boldsymbol{M_0}$. Zdroj týchto údajov predstavuje editor matíc, v ktorom je možné zadať:

- hodnoty váh vstupných hrán prechodov – matica ***Pre***
- hodnoty váh výstupných hrán prechodov – matica ***Post***
- počiatočné značenie Petriho siete – vektor $\boldsymbol{M_0}$
- kapacity jednotlivých miest

Okrem editora matíc je možné potrebné údaje načítať z textového súboru v predvolenom formáte. Zadané údaje je takisto možné do textového súboru aj uložiť.

B. Editor údajov potrebných pre analýzu časovanej PS alebo PS s priradenými nákladmi

Ďalšie vstupy od používateľa sú po načítaní Petriho siete aj dáta, ktoré sú potrebné pre analyzovanie výkonnosti výrobnej linky a štatistické výpočty. Na tento účel bolo vytvorený editor pre úpravu týchto dát. Obsahuje tieto funkcie:

- zadávanie časov a nákladov pre jednotlivé miesta P-časovanej PS
- zadávanie časov a nákladov pre jednotlivé prechody T-časovanej PS
- výpočet času alebo nákladov potrebných na výrobu 1 výrobku alebo po konkrétny stav v sieti
- zadávanie požadovaných hodnôt, pre ktoré je možné vykonať analýzu PS a zistiť tak predpokladanú výrobu

C. Okno výsledkov analýzy

Pre zobrazenie výsledkov analýzy bolo vytvorené okno, ktoré zobrazuje dané informácie v jednotlivých tabuľkách. V okne sa môže vyskytnúť maximálne 5 tabuliek:

- tabuľka s výsledkom času a nákladov potrebných na výrobu 1 výrobku
- tabuľka pre používateľom zadaný čas, ktorá zobrazuje informácie o celkových nákladoch a počte výrobkov pri danej výrobe, a čase, ktorý nebude pri výrobe využitý
- tabuľka pre používateľom zadané náklady, ktorá zobrazuje informácie o celkovom čase a počte výrobkov pri danej výrobe, a nákladoch, ktoré nebudú pri výrobe využité
- tabuľka pre používateľom zadaný počet výrobkov, ktorá zobrazuje informácie o celkových nákladoch a čase potrebnom pri danej výrobe, a čase, ktorý nebude pri výrobe využitý
- tabuľka pre zobrazenie času potrebného na údržbu výrobnej linky pre zadaný požadovaný hodinový výkon linky

D. Globálne štatistiky pre miesta a prechody PS

Z okna analýzy je možné zobraziť štatistiky pre miesta a prechody PS, ktoré vznikli pri simulácii PS. Obsahuje informácie o počte značiek, ktoré odišli a prišli z/do jednotlivých miest, priemernom čase medzi dvoma odchodmi/príchodmi značiek z/do jednotlivých miest, hodnoty priemerného času čakania značky v jednotlivých miestach a počet prepálení jednotlivých prechodov a priemerný čas medzi dvoma úspešnými prepáleniami daného prechodu.

Údaje je možné exportovať do Excel dokumentu.

E. Výpočet prepaľovacej sekvencie

Na zistenie prepaľovacej sekvencie, ktorú je potrebné použiť pri transformácii PS zo stavu M_0 do stavu M_X bolo vytvorené okno pre tento výpočet. Vyžaduje poznanie požadovaného stavu, pre ktorý chceme určiť sekvenciu prepálenia jednotlivých prechodov. Tento stav je možné vyčítať zo zoznamu dosiahnuteľných stavov Petriho siete.

F. Dosiahnuteľné stavy

Pre efektívnu simuláciu a analýzu PS je potrebné poznať všetky možné stavy, do ktorých sa daná PS dostane. V okne dosiahnuteľných stavov PS je zoznam stavov dosiahnuteľných zo stavu M_0.

G. Náhľad zadaných údajov

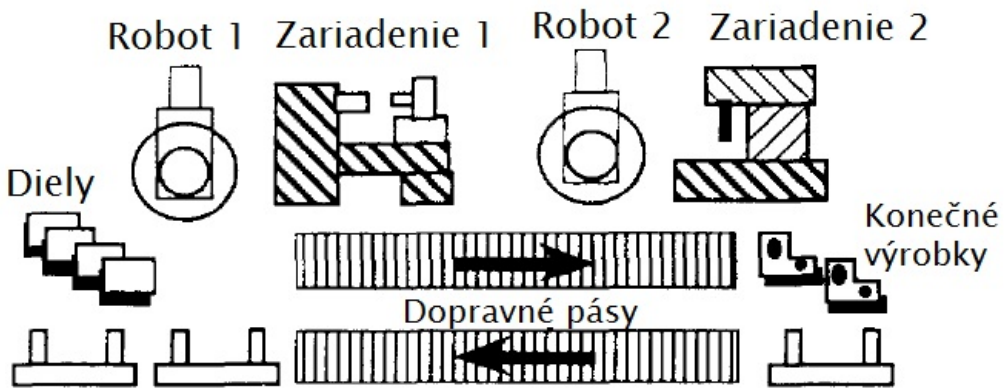
Pre prehľad zadaných údajov, čiže údajov z editora matíc bolo vytvorené prehľadné okno, ktoré slúži na prístup k daným maticiam pri práci s programom.

V. PRÍKLAD VÝROBNEJ LINKY

Na zadanom modeli výrobného procesu bude ukázaná funkčnosť a možnosti aplikácie.

A. Popis výrobnej linky

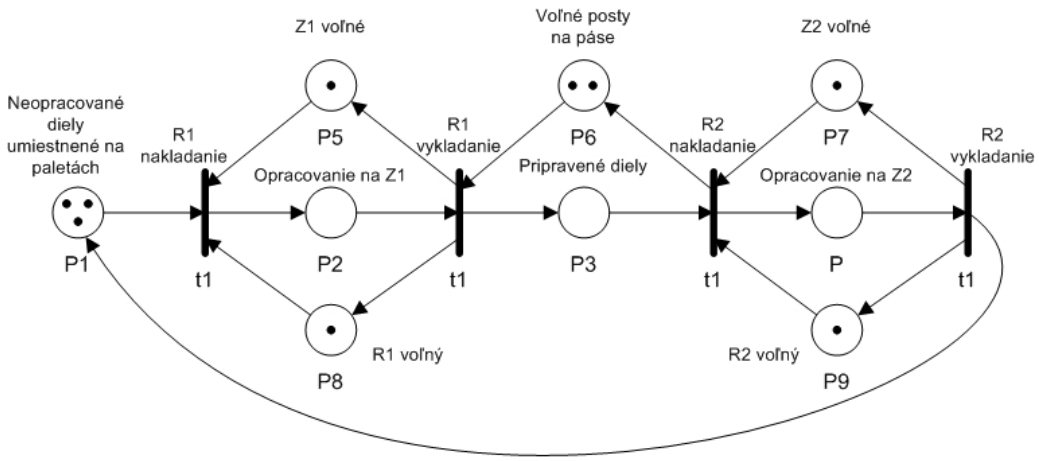
Majme výrobnú linku skladajúcu sa z týchto zariadení a častí:



Obr. 1 Zariadenia a časti výrobnej linky [3]

Výrobná linka pozostáva z dvoch zariadení na montáž dielcov, dvoch manipulátorov a dvoch dopravných pásov. Každé montážne zariadenie je obsluhované jedným ramenom, ktoré vykonáva ukladanie a vyberanie dielov. Jeden dopravný pás sa používa na prepravu dielov, najviac dva naraz. Po druhom páse sa prepravujú prázdne palety. V tomto systéme sú k dispozícii 3 palety. Každý diel je montovaný a opracovávaný na oboch strojoch Z1 a Z2 v tomto poradí. Proces, ktorý sa vykonáva na zariadení Z1 trvá 10 sekúnd a je naňho potrebné vynaložiť náklady 5€, na druhom zariadení to je 16 sekúnd a náklady sú 6€. [3]

Výrobný proces, ktorý sa vykonáva prostredníctvom tejto linky je možné namodelovať pomocou Petriho siete. Na Obr. 2 sa nachádza zjednodušený model takéhoto procesu.



Obr. 2 Model výrobnej linky

B. Výstupy z aplikácie

Z daného grafu a z popisu výrobnej linky môžme vyčítať matice ***Pre***, ***Post***, vektor M_0, časy a náklady potrebné na realizáciu jednotlivých procesov. Tieto údaje je možné zadať do aplikácie na ďalšie spracovanie. Výstup zadaných matíc je na Obr. 3 a Obr. 4.

Pre

	t1	t2	t3	t4
P1	1	0	0	0
P2	0	1	0	0
P3	0	0	1	0
P4	0	0	0	1
P5	1	0	0	0
P6	0	1	0	0
P7	0	0	1	0
P8	1	0	0	0
P9	0	0	1	0

Post

	t1	t2	t3	t4
P1	0	0	0	1
P2	1	0	0	0
P3	0	1	0	0
P4	0	0	1	0
P5	0	1	0	0
P6	0	0	1	0
P7	0	0	0	1
P8	0	1	0	0
P9	0	0	0	1

M0

P1	3
P2	0
P3	0
P4	0
P5	1
P6	2
P7	1
P8	1
P9	1

Obr. 3 Matice ***Pre***, ***Post*** a vektor M_0

Incidenčná matica

	t1	t2	t3	t4
P1	-1	0	0	1
P2	1	-1	0	0
P3	0	1	-1	0
P4	0	0	1	-1
P5	-1	1	0	0
P6	0	-1	1	0
P7	0	0	-1	1
P8	-1	1	0	0
P9	0	0	-1	1

Obr. 4 Incidenčná matica

M0 : 3 0 0 0 1 2 1 1 1
M1 : 2 1 0 0 0 2 1 0 1
M2 : 2 0 1 0 1 1 1 1 1
M3 : 1 1 0 1 0 2 0 0 0
M4 : 2 0 1 0 1 1 1 1 1
M5 : 1 1 0 1 0 2 0 0 0
M6 : 2 0 1 0 1 1 1 1 1
M7 : 1 1 0 1 0 2 0 0 0
M8 : 2 0 1 0 1 1 1 1 1
M9 : 1 1 0 1 0 2 0 0 0

< >

Obr. 5 Dosiahnuteľné stavy zo stavu M_0

Zadané údaje je možné použiť pri výpočte prepaľovacej sekvencie na základe niektorého zo stavov dosiahnuteľných zo stavu M_0 (Obr. 5). Pri zadaní stavu M_3 je možné zistiť, ktoré prechody budú prepálené, pokiaľ sa sieť nedostane do zadaného stavu M_3. Na základe vzorca:

$$C^+ * \sigma = \Delta M,$$

a Penroseovho algoritmu pre výpočet pseudoinverznej matice bude výstup programu

σ3: 2t1, t2, t3

kde $\Delta M = M_d - M_0$ a $\sigma = \sum_{k=1}^{d} u_k$. Vektor u predstavuje postupnosť prepálení a matica C^+ je pseudoinverznou k incidenčnej matici C.

Znamená to, že PS potrebuje na transformáciu zo stavu M_0 do stavu M_3 prepáliť 2-krát prechod t1 a raz prepáliť prechody t2 a t3.

Hlavná časť aplikácie nám umožní analyzovať danú PS. Z popisu výrobnej linky zistíme, že ide o P-časovanú PS a miesta P2 a P4 majú priradené časy 10 a 16 sekúnd a náklady 5€ a 6€. Po zadaní týchto údajov je ešte možné PS analyzovať z hľadiska konkrétneho času a nákladov, počtu výrobkov a je možné vypočítať čas údržby počas x-hodinovej zmeny.

Vstupy:

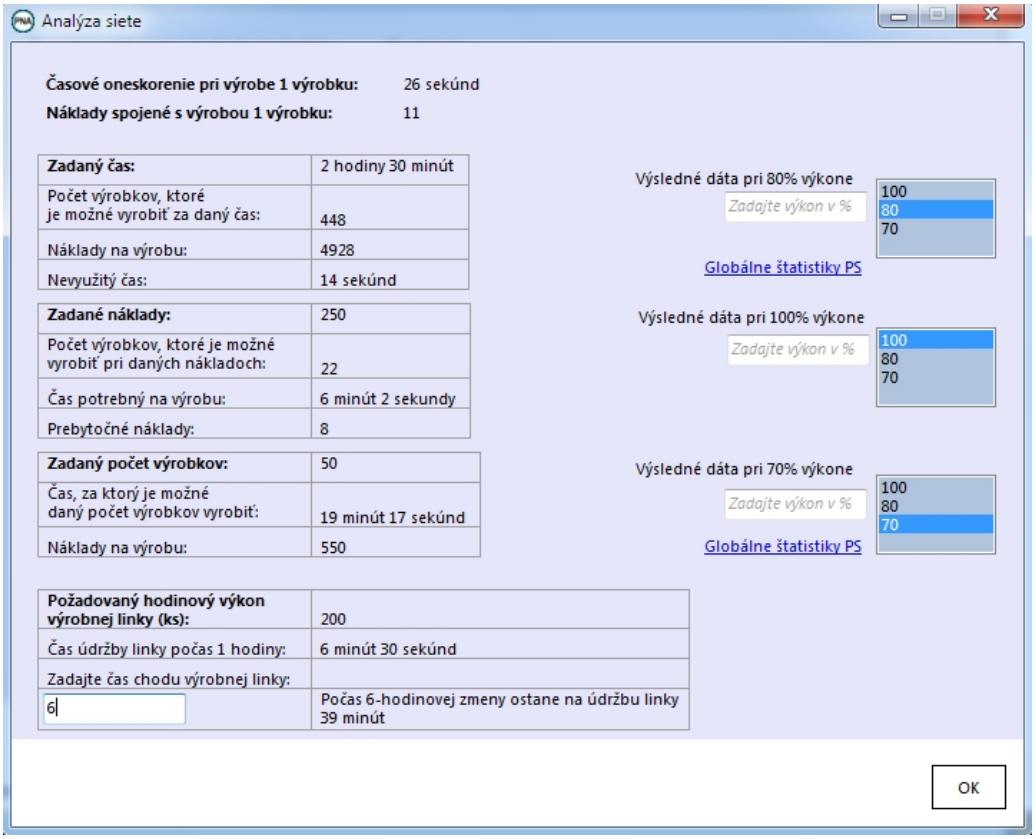
čas – 2 hodiny, 30 minút
počet výrobkov – 50
náklady - 250€
hodinový výkon linky – 200

Výstupy:

Na Obr. 6 sú zobrazené výsledné údaje.

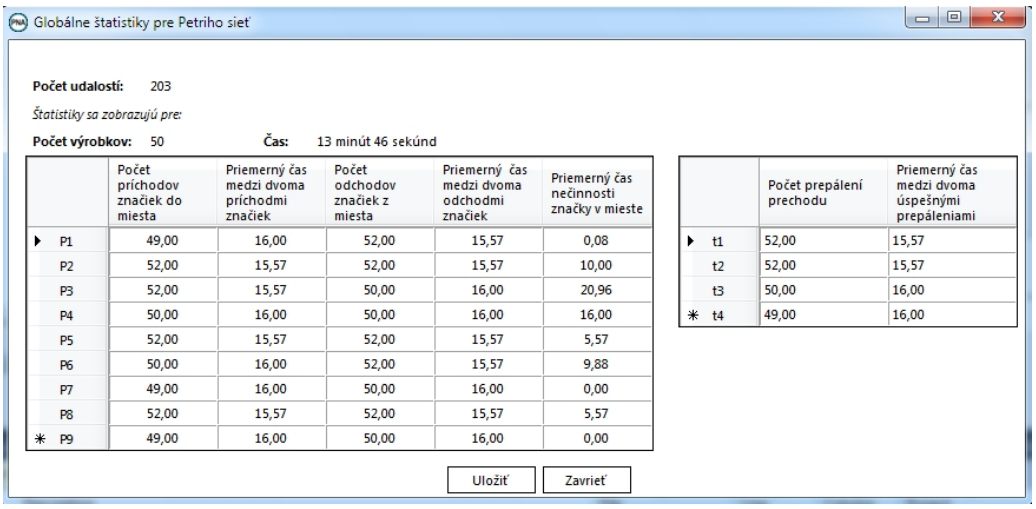
- Pre zadaný čas je možné vyrobiť reálne 392 až 448 výrobkov, čo je oproti ideálnemu výkonu výrobnej linky zníženie o takmer o 100 až 160 výrobkov. Podľa [2] je ideálny hodinový výkon 100%, pričom reálny je znížený o 20 až 30%, u niektorých systémov aj viac. Je to z dôvodu určitých nárokov na obsluhu zariadenia. Obsluha zariadenia môže predstavovať napr. odstraňovanie chýb, výmenu chýbajúceho materiálu na paletách, údržbu výrobnej linky a pod.
- Pre zadané náklady je zobrazená predpokladaná výroba pri 100% ideálnom výkone. Reálne sa za 250€ dá vyrobiť 22 výrobkov, ale pri zvýšenom čase o 106 až 228 sekúnd.
- Daný počet výrobkov rovný 50 je možné reálne vyrobiť, ale takisto sa zvyšuje čas oproti ideálnej výrobe o takmer 3 a pol až 6 minút.
- Nakoniec je pre hodinový výkon linky vypočítaný čas, ktorý ostáva na údržbu linky, čo je 6 a pol minúty na 1 hodinu behu linky. Pri 6 hodinovej zmene je teda možné na údržbu linky obetovať 39 minút.

V editori údajov pre analýzu je ešte možné použiť zadané údaje na vypočítanie času výroby a nákladov na ňu po nejaký špecifický stav. Ak napr. zadáme stav $\boldsymbol{M_5}$: [0 1 1 1 0 1 0 0 0] z okna dosiahnuteľných stavov, ale pre časovanú sieť, po analýze zistíme, že ak sa chceme dopracovať k tomuto stavu, bude to trvať 20 sekúnd a stáť 10€.



Obr. 6 Výsledky analýzy

Aplikácia ešte umožňuje zobraziť štatistiky pre miesta a prechody pri 100% výkone linky pre zadaný čas alebo počet výrobkov. Výsledné údaje sú na Obr. 7.



Obr. 7 Globálne štatistiky

VI. ZÁVER

Výrobné procesy sú zložitým problémom a správny návrh ich riadenia a návrh výrobných liniek je veľmi dôležitý z hľadiska efektívnej výroby a optimalizácie výrobných nákladov a času potrebného na výrobu. Táto aplikácia môže napomôcť v odsimulovaní daného výrobného procesu a prípadnom jeho doladení a vylepšení.

V tomto štádiu podporuje hlavne časované Petriho siete, ale v budúcnosti je priestor na rozšírenie aplikácie aj o iné typy Petriho sietí a takisto o ďalšie funkcie. Je možné pridať grafický editor a zobrazovanie ďalších štatistík a výpočtov pri analýze PS. Takisto by bolo vhodné optimalizovať algoritmy, ktoré by zahŕňali čo najväčší počet prípadov, ktoré môžu pri modelovaní pomocou Petriho sietí nastať.

POĎAKOVANIE

Táto práca bola vytvorená realizáciou projektu Rozvoj Centra informačných a komunikačných technológií pre znalostné systémy (kód ITMS projektu: 26220120030) na základe podpory operačného programu Výskum a vývoj financovaného z Európskeho fondu regionálneho rozvoja.

Táto práca bola podporená z vedeckého projektu Vega č. 1/0286/11 Grantovej agentúry SR pod názvom „Dynamické hybridné architektúry v multiagentových sieťových riadiacich systémoch“.

LITERATÚRA

[1] MURATA, T. 1989. Petri nets: Properties, analysis and applications. In Proceedings of the IEEE, Vol.77, No.4, s 541 – 580

[2] MADARÁSZ, L. 2006. Integračné aspekty tvorby a prevádzky systémov CIM : 2.vydanie. Košice : Elfa, s.r.o., 2006. 380 s. ISBN 80-86-043-2

[3] ZURAWSKI, Richard – ZHOU, MengChu: Petri Nets and Industrial Applications: A Tutorial. IEEE Transactions on Industrial Electronics [online]. IEEE Industrial Electronics Society, Vol.41, No.6, 1994. Dostupné na internete: < http://class.ee.iastate.edu/cpre545/refs/PetriNets1.pdf>. ISSN 0278-0046.

[4] COMAN, Daniela et al.: Manufacturing System Modeling Using Petri Nets. International Conference on Economic Engineering and Manufacturing Systems [online]. Recent, Vol.10 (2009), No.3 (27). Dostupné na internete: < http://www.recentonline.ro/027/COMAN_Daniela.pdf>.

Vyhľadávanie dokumentov s použitím formálnej konceptuálnej analýzy

[1]*Martin SARNOVSKÝ,* [1]*Marek BALUCHA,* [2]*Peter BUTKA*

[1]Katedra kybernetiky a umelej inteligencie, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika
[2]Centrum pre informačné technológie, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

[1]martin.sarnovsky@tuke.sk, [1]marek.balucha@student.tuke.sk, [2]peter.butka@tuke.sk

***Abstrakt* — Táto práca sa zaoberá problematikou vyhľadávania webových stránok použitím prístupu formálnej konceptuálnej analýzy (FCA). Podáva teoretický pohľad na problematiku vyhľadávania informácií, predspracovania textových dokumentov a modelov založených na formálnej konceptuálnej analýze. Hlavným cieľom tejto práce je vybudovanie modelu založeného na FCA a jeho prehľadávanie, ako aj niekoľko príkladov testovania vyhľadávania pomocou tohto modelu.**

***Kľúčové slová* — formálna konceptuálna analýza, vyhľadávanie, spracovanie textov, konceptuálne modely**

I. Úvod

Informácie majú v dnešnom svete významnú úlohu. Pomáhajú subjektom, ktoré ich požadujú, uskutočňovať správne rozhodnutia v danom čase. Z tohto dôvodu sa veľa firiem snaží získavať informácie, analyzovať ich, uchovávať a spravovať pomocou rôznych nástrojov a databáz, umožniť k nim prístup všetkým požadovaným osobám, ktoré ich môžu využívať na efektívne a bezproblémové uskutočňovanie podnikových procesov.

Predmetom záujmu tejto práce je vyhľadávanie webových stránok a iných textových zdrojov použitím prístupu formálnej konceptuálnej analýzy (FCA). Vyhľadávanie informácií (IR z angl. Information Retrieval) ako vedná disciplína sa zaoberá reprezentovaním, uchovávaním, organizáciou informácií a prístupom k týmto informačným položkám. Reprezentácia a organizácia informácií by mali umožniť používateľovi jednoduchý prístup k informáciám, o ktoré ma záujem. Nanešťastie charakterizácia používateľskej informačnej potreby nie je jednoduchý problém. Hlavnými cieľmi vyhľadávania informácií sú indexácia textu a hľadanie užitočných dokumentov vo vybratej kolekcii dokumentov. Výskum v oblasti IR sa zaoberá predovšetkým tvorbou IR modelov, klasifikáciou a kategorizáciou dokumentov, architektúrou systémov, používateľským rozhraním, vizualizáciou dát, či filtrovaním dát [1].

Základnými pojmami v oblasti FCA sú formálny kontext a koncept [4]. Všetky formálne koncepty a ich vzťahy sa graficky znázorňujú konceptovým zväzom. Vstupné dáta pre testovanie boli tvorené z webových stránok z domény www.goodreads.com uložených v XML formáte. Pre prácu s dokumentmi bola využitá softvérová knižnica JBowl [2]. Na vybudovanie modelu založeného na formálnej konceptuálnej analýze sa použil inkrementálny algoritmus pre tvorbu zovšeobecnených jednostranných konceptových zväzov [3].

Začiatok práce bude venovaný základným pojmom v rámci predspracovania a modelov množín dokumentov. V nasledujúcej kapitole bude popísaný použitý model na báze FCA, spolu s popisom niektorých implementačných detailov. Následne si predstavíme návrh spôsobu prehľadávania vybudovaného modelu založeného na FCA, vytvorenú webovú aplikáciu určenú na vyhľadávanie webových stránok použitím prístupu FCA a jej otestovanie a vyhodnotenie prostredníctvom niekoľkých experimentov.

II. Predspracovanie a modely množín textových dokumentov

Predspracovaním nazývame proces transformácie neštruktúrovaných informácií do

štruktúrovanej reprezentácie [5]. Pod takouto štruktúrovanou reprezentáciou môžeme chápať napr. maticu dokument - term. Entitami tejto matice sú jednotlivé textové dokumenty a atribútmi sú váhy výskytov jednotlivých termov v danom dokumente, ktorých hodnota závisí od použitého modelu reprezentácie textových dokumentov. Predspracovanie textových dokumentov môžeme chápať ako proces, ktorého vstupom sú elektronické textové dokumenty a výstupom je vektorová reprezentácia dokumentu. Podľa [6] jednotlivé kroky zahŕňajú:

- Konverzia na čistý text - čistý text predstavuje sekvenciu alfanumerických, interpunkčných, oddeľovacích grafických znakov a niektorých špeciálnych symbolov (napr. %, &, *, atď.). Z rôznych formátov textových dokumentov môžeme získať čistý text tak, že z nich odstránime formátovacie znaky, netextové informácie atď. Pri získavaní čistého textu je potrebné brať v úvahu kódovanie, odporúča sa používať UTF8. Informácie, ktoré majú istú vypovedaciu hodnotu, ale pri konverzií na čistý text sa strácajú (metainformácie), môžeme zachovať (napr. archiváciou pôvodného dokumentu s referenciou na príslušný čistý text, XML notáciou atď.) a v prípade potreby k nim pristúpiť a použiť ich.
- Tokenizácia a segmentácia - v tomto kroku predspracovania sa čistý text rozčlení na elementárne textové reťazce na základe nejakého oddeľovača, napr. medzera. Následne je potrebné jednotlivé segmenty previesť na tokeny, čo sú slová, resp. slovné spojenia s určitým významom. Tokeny sa vytvárajú za pomoci slovníkov prípustných tvarov v kombinácií s istými pravidlovými systémami. Tokenom môžu byť napr. ustálené slovné spojenia, zložené tvary a slova, rôzne kombinácie čísel, písmen, interpunkčných znamienok a znakov, ktorých kombinácia má určitý špecifický význam, napr. konkrétny dátum viažuci sa na istú významnú udalosť, verzie softvérových produktov, názvy hudobných kapiel, webových sídel atď.
- Lematizácia, morfologická analýza - lematizáciou, čiže prevodom slova na jeho základný tvár (lemu) významne zredukujeme počet potencionálnych termov. A to z tohto dôvodu, že význam koreňa slova je vždy rovnaký, ale iba za predpokladu významovej jednoznačnosti danej lemy. V takomto prípade je vhodne problém riešiť morfologickou analýzou, čiže zaoberať sa aj slovnými druhmi a gramatickými kategóriami lem. Na izoláciu koreňa slova sa pre anglické texty najčastejšie využíva Porterov algoritmus.
- Eliminácia, váhovanie, normovanie - berúc v úvahu fakt, že termy by mali čo najpresnejšie vystihovať obsah daného textu, je vhodne uchovávať iba plnovýznamové slova. Z doteraz získanej množiny potencionálnych termov je možné identifikovať a následne odstrániť neplnovýznamové slova, napr. pomocou slovníka, ktorý obsahuje zoznam slov, ktoré nie sú vhodnými adeptmi na post termu. Ďalším, jednoduchším, ale menej uspokojivým spôsobom je odstrániť iba tie slová, ktoré z hľadiska frekvencie v globálnej množine dokumentov patria do intervalu <n/100, n/10>, pričom n je počet dokumentov. Váhovaním vypočítame pre každý term v rámci daného dokumentu váhu, t. j. mieru dôležitosti daného termu pre daný dokument. Výpočet závisí od použitého modelu reprezentácie textových dokumentov. Následne sa vypočítané váhy normujú, tzn., že hodnoty váh sa prevedú do hodnôt stanoveného intervalu napr. <0, 1>.

Modelom reprezentácie textových dokumentov nazývame takú formalizáciu dokumentu, resp. jeho textu v prirodzenom jazyku, ktorá nejakým spôsobom vyjadruje obsah tohto textu a zároveň umožňuje ich efektívne spracovanie metódami dolovania v textoch. Procesom predspracovania textových dokumentov získame pre každý dokument zredukovaný vektor lematizovaných termov, ktoré najviac vyjadrujú obsahovú stránku daného dokumentu. Sú to hlavne podstatné mená, prídavne mena a slovesa. Na základe získaných vektorov termov môžeme vytvoriť maticu dokument – term, ktorá predstavuje množinu vstupných dát (v podobe boolean hodnôt, rôznych váh termov atď.) pre jednotlivé modely reprezentácie súboru textových dokumentov. Proces transformácie váhovaných a normovaných termov do relačnej štruktúry matice dokument - term sa nazýva indexacia. Matica dokument – term predstavuje vhodnú formu reprezentácie dokumentov, ale pri veľkom množstve dokumentov a termov je lepšou voľbou invertovaný index, ktorý obsahuje informácie o termoch vyskytujúcich sa v dokumentoch (nenulové hodnoty v matici dokument - term), o ich príslušnosti k jednotlivým dokumentom a pripadne aj o pozíciách ich výskytov.

Boolovský model predstavuje najjednoduchší model reprezentácie textových dokumentov.

Obsahuje iba informáciu o tom, či sa daný term v dokumente vyskytuje alebo nie. Term môže nadobúdať iba binárne hodnoty 0 (v prípade ak sa v dokumente nevyskytuje) a 1 (ak sa v dokumente vyskytuje aspoň raz). Booleovský model vyhľadávania informácií je založený na teórií množín a booleovskej algebre [1]. Dopyt môže obsahovať logické operácie konjunkcie (AND), disjunkcie (OR) a negácie (NOT). Booleovský model nesie so sebou množstvo problémov. Medzi najpodstatnejší patri nutnosť presnej zhody výskytu termov v dopyte a v hľadanom dokumente, čoho dôsledkom je buď príliš veľký (pri operáciách disjunkcie), alebo príliš malý (pri operáciách konjunkcie) počet vyhľadaných dokumentov. Získané dokumenty nie je možne zoradiť podľa stupňa relevancie, čo v konečnom dôsledku sťažuje nájdenie požadovaného dokumentu.

Vektorový model odstraňuje mnohé nevýhody boolovského modelu. Neuvažuje už iba o výskyte daného termu v dokumente, ale berie v úvahu viacero faktorov. Zohľadňuje dôležitosť daného termu, tzn. akou mierou daný term reprezentuje dokument, v ktorom sa nachádza. Dôležitosť daného termu sa zohľadňuje pomocou rôznych váh. Najjednoduchšie sa váhy vyjadrujú na základe frekvencie daného termu v dokumente (term frequency - *tf*). Čím častejšie sa daný term v danom dokumente vyskytuje, tým ma väčšiu váhu, čiže je lepším identifikátorom dokumentu, v ktorom sa nachádza. Avšak relevancia termu nerastie proporcionálne (lineárne) s frekvenciou jeho výskytu v texte, to znamená, že term, ktorý sa v dokumente vyskytuje n – krát, nebude aj n – krát relevantnejší. Z tohto dôvodu je váha kladnej termovej frekvencie rovná:

$$\mathrm{w}_{\mathrm{t,d}} = 1 + \log_{10} tf_{t,d}$$

Najlepšími identifikátormi dokumentu sú termy, ktoré sa v danom terme vyskytujú často, ale v ostatných dokumentoch sa nevyskytujú, resp. sa vyskytujú iba sporadicky. Z tohto dôvodu je nutné brať v úvahu aj frekvenciu termu v celom korpuse dokumentov, tzv. dokumentovú frekvenciu *df*. Pri dokumentovej frekvencií je potrebné poznať celkový počet dokumentov v korpuse *N*. Dokumentovú frekvenciu je taktiež potrebné logaritmovať, čím získame tzv. inverznú dokumentovú frekvenciu pre term *t*

$$idf_t = \log_{10}\left(\frac{N}{df_t}\right)$$

Kombináciou termovej a dokumentovej frekvencie získame váhovaciu schému tf-idf:

$$w_{t,d} = tf_{t,d} \times idf_t$$

Pri porovnávaní podobnosti vektorov váh je potrebné hodnoty normovať, tzn. zjednotiť dĺžky vektorov. Najčastejšie sa používa tzv. kosínusová miera podobnosti, ktorá sa rovná skalárnemu súčinu zložiek vektorov, pričom súčin je normalizovaný. Podobnosť vektorov dopytu q a dokumentu d získame nasledovne:

$$\cos(\vec{q},\vec{d}) = \frac{\sum_{i=1}^{|T|} q_i d_i}{\sqrt{\sum_{i=1}^{|T|} q_i^{\,2}}\sqrt{\sum_{i=1}^{|T|} d_i^{\,2}}}$$

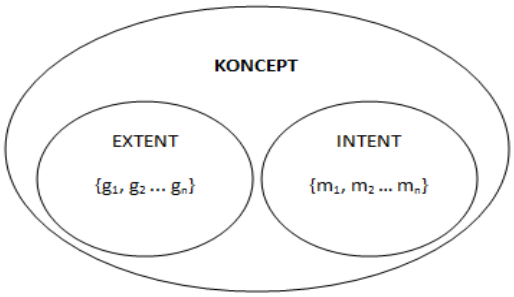
Čim viac sa výsledná hodnota približuje k hodnote 1, tým viac je dokument *d* relevantnejší k dopytu *q*. Tento model nám umožňuje usporiadať vyhľadané dokumenty podľa stupňa podobnosti k danému dopytu.

III. KONCEPTUÁLNY MODEL NA BÁZE FCA

V tejto kapitole si predstavíme modely, ktoré využívajú konceptovú analýzu dát. Každý koncept sa skladá z objektovej (zvanej extent) a atribútovej zložky (zvanej intent). Modely, ktoré si predstavíme v tejto kapitole sú založené na teórií zväzov. Čiastočným alebo úplným usporiadaním konceptov získame štruktúru zvanú konceptový zväz. Konceptové zväzy sú zaujímavé aj z hľadiska vizualizácie dát. Vizualizácia konceptového zväzu pomocou Hasseovho diagramu predstavuje štruktúru, v ktorej sa človek dokáže pomerne ľahko orientovať.

FCA navrhol Rudolf Wille v roku 1982. Najmä v posledných desiatich rokoch bolo FCA široko uplatňované v mnohých vedných disciplínach ako napríklad lingvistika, umelá inteligencia, softvérové inžinierstvo a vyhľadávanie informácii atď. FCA je metóda orientovaná na človeka, k štruktúrovaniu a analýze dát, reprezentácií znalosti a správe informácií. V FCA sú dáta (dokumenty) spolu s ich kontextom (termami) štruktúrované do

formálnych konceptov, ktoré takto tvoria konceptový zväz, ktorý je usporiadaný na základe tzv. subkoncept – superkoncept vzťahu. Všetky dokumenty v danom zhluku, ktoré predstavuje formálny koncept (Obr. 1) sa označujú ako extenty a všetky atribúty ako intenty. Konceptový zväz predstavuje graf pozostávajúci z uzlov a prepojení.



Obr. 1 Komponenty konceptu v rámci FCA.

Klasický rámec FCA rozširuje zovšeobecnený jednostranne fuzzy konceptový zväz, tzv. GOSCL, ktorý na rozdiel od FCA dokáže pracovať s nehomogénnymi typmi atribútov. Následne si podľa [3] definujeme zovšeobecnený jednostranný konceptový zväz:

Usporiadanej množine štyroch prvkov c = (*B*, *A*, *L*, *R*) hovoríme zovšeobecnený jednostranný formálny kontext, ak spĺňa nasledujúce podmienky:

1. *B* je neprázdna množina objektov a *A* je neprázdna množina atribútov.
2. *L*: *A* → *CL* je zobrazenie z množiny atribútov ku skupine všetkých kompletných zväzov. Odteraz, pre každý atribút *a*, *L(a)* uvádzame štruktúru pravdivostných hodnôt pre atribút *a*.
3. *R* predstavuje zovšeobecnený vzťah incidencie, t. j., *R* (*b*,*a*) ∈ L(*a*) pre každé *b* ∈ *B* a *a* ∈ *A*. Teda *R*(*b*, *a*) reprezentuje stav zo štruktúry *L(a)* v ktorej element *b* ∈ *B* má atribút *a*.

Teraz definujeme základné poznatky o zovšeobecnených jednostranných konceptových zväzoch. Nech (*B*, *A*, *L*, *R*) je zovšeobecneným jednostranným formálnym kontextom. Potom definujeme dvojicu zobrazení $\uparrow: 2^B \to \prod_{a\in A} L(a)$ and $\downarrow: \prod_{a\in A} L(a) \to 2^B$ nasledovne:

$$\uparrow(X)(a) = \inf_{b\in X}(R(b,a))$$

$$\downarrow(g) = \{\, b \in B:\ \forall a \in A, g(a) \le R(b,a)\}$$

Nech (*B*, *A*, *L*, *R*) je zovšeobecneným jednostranným kontextom. Potom dvojica (↑,↓) tvorí Galoisove spojenie medzi 2^B a $\prod_{a\in A} L(a)$.

Teraz je možné definovať zovšeobecnený jednostranný konceptový zväz. Pre formálny kontext (*B*, *A*, *L*, *R*) označme C (*B*, *A*, *L*, *R*) množinou všetkých dvojíc (*X*, *h*), kde *X* je podmnožinou *B*, $b \in \prod_{a\in A} L(a)$, ktoré spĺňajú ↑(*X*)=*h* a ↓(*h*) = *X*. Potom množina *X* je zvyčajne označovaná ako extent a *h* ako intent konceptu (*X*, *h*).

Ďalej definujeme čiastočné usporiadanie C (G, M, Γ, R) nasledovne: $(X_1, h_1) \le (X_2, h_2)$ vtedy a len vtedy, ak $X_1 \subseteq X_2$ a $h_1 \ge h_2$. Nech (*B*, *A*, *L*, *R*) je zovšeobecneným jednostranným formálnym kontextom. Potom C (*B*, *A*, *L*, *R*) s čiastočným usporiadaním definovaným vyššie vytvára kompletný zväz, tzv. zovšeobecnený jednostranný konceptový zväz.

Implementácia modelu je postavená na projekt GOSCL [3], vytvoreného v objektovo orientovanom programovacom jazyku Java, ktorý umožňuje vytvárať konceptové zväzy, pričom vstupné atribúty môžu byť rôznorodé. V rámci tohto softvéru je možné vytvárať konceptové zväzy, používať vytvorené štruktúry a zobrazovať ich pomocou generovaných obrázkov. Algoritmy a postupy odtiaľto boli integrované do balíka JBOWL a použité v ďalšej časti práce.

IV. VYHĽADÁVANIE S POUŽITÍM MODELU NA BÁZE FCA

Keďže hlavným cieľom bol návrh, implementácia a otestovanie IR (Information Retrieval) systému pre vyhľadávanie informácií z webových stránok použitím prístupu formálnej konceptuálnej analýzy, model vytvorený podľa postupu prezentovaného v predchádzajúcej kapitole bol použitý pre vyhľadávanie. Na predspracovanie, indexáciu vstupných dát a tvorbu modelu bola využitá javovská softvérová knižnica JBowl (Java Bag-of-words Library), ktorá je

bližšie popísaná v [2]. Pre potreby zostavenia požadovaného modelu bol z projektu GOSCL [3] získaný inkrementálny algoritmus pre tvorbu zovšeobecnených jednostranných konceptových zväzov a implementovaný do knižnice JBowl tak, aby využíval jeho štruktúry. Webová aplikácia na vyhľadávanie webových stránok využíva technológiu JSP (Java Server Pages).

Vstupné dáta predstavujú webové stránky obsahujúce recenzie kníh v anglickom jazyku, ktoré boli získane z jednej domény: www.goodreads.com. Knižnica JBowl používa špecifickú formu vstupu v XML formáte. Z tohto dôvodu bolo potrebné webové stránky previesť nasledujúceho tvaru:

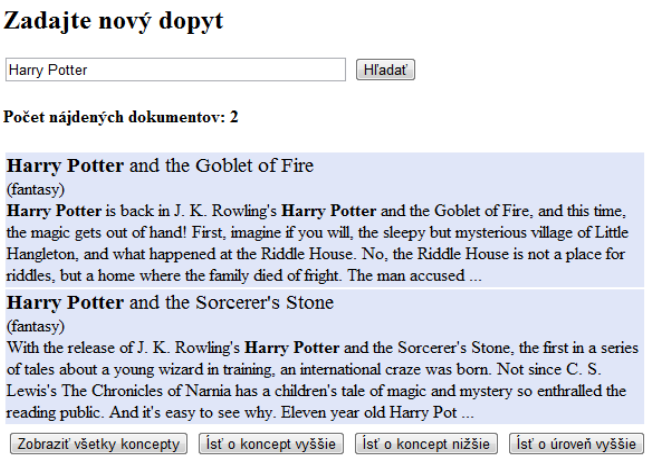
```
<dataset>
    <document id="" name="">
      <author></author>
      <title></title>
      <category></category>
      <rating></rating>
      <text></text>
    </document>
</dataset>
```

Každý element document reprezentuje jednu webovú stánku. Množina dát obsahuje celkovo 50 dokumentov. Každý dokument obsahuje informácie typu: autor, resp. autori knižného titulu, názov knižného titulu, žánrovú kategóriu, rating knižného titulu a samotný text recenzie.

Vybudovalo sa sedem modelov zovšeobecnených jednostranných konceptových zväzov. Prvý bol vybudovaný zo vstupných dát, ktoré boli získané z procesu predspracovania. Zvyšných šesť boli vybudovávané zo vstupných dát, ktoré okrem procesu predspracovania boli orezané o termy, ktoré mali maximálnu hornú hraničnú frekvenciu nastavenú na hodnoty: 2, 3, 4, 5, 6 a 7. Samozrejme, počet konceptov v konceptovom zväze rástol v závislosti od počtu termov, a to od 92 konceptov pre prípad 2800 termov, až po 154000 konceptov pre prípad 4528 termov.

Vytvorený model prestavuje štruktúru, z ktorej ak chceme získať nejakú informáciu, tak musíme vedieť k nej pristúpiť. Pristupovať k potrebným údajom môžeme pomocou dostupných metód. V našom prípade potrebujeme získať taký koncept, resp. koncepty, ktorých extent obsahuje objekty, ktoré sú pre nás v danom okamihu zaujímavé. Následne si predstavíme spôsob vyhľadania, ktorý je založený na podobnosti vektora dopytu s intentom. Intent je reprezentovaný vektorom normovaných váh termov. Získanie informácii o miere podobnosti medzi vektorom reprezentujúcim dopyt a intentom príslušného konceptu v rámci cyklu sa doshaovalo pomocou kosínusovej metriky.

Webová aplikácia pre zadávanie dopytu a zobrazenie vyhľadávania bola vytvorená v prostredí NetBeans a využíva technológiu JSP (Java Server Pages). Vo vytvorenej webovej aplikácií sa vytvorili JSP súbory Query (obsahuje formulár, kde používateľ vzhľadom na svoju informačnú potrebu môže zadať svoj dopyt v podobe kľúčových slov) a Retrieval (obsahuje implementáciu vyhľadávania dokumentov z konceptov v rámci konceptového zväzu). Následne bol vyhľadaný k dopytu najpodobnejší koncept a dokumenty, ktoré sa vyskytovali v jeho extente, boli zobrazené ako výsledok vyhľadávania. Príklad výsledku vyhľadávania prostredníctvom tohto súboru a aj základný užívateľský pohľad na funkčnosť (pohyb v rámci konceptovej štruktúry) je možné vidieť na Obr. 2.



Obr. 2 Zobrazený výstup vyhľadávania prostredníctvom súboru retrieval.jsp

Výstupom zobrazeným na obrazovke výstupu boli všetky neredundantné dokumenty všetkých vyhľadaných konceptov. Výstup bol používateľovi zobrazený obdobne s popisom vyššie. Ak používateľ nebol spokojný s výsledkom vyhľadávania, mohol zadať nové kľúčové slova alebo sa skúsiť pozrieť od najpodobnejšieho konceptu o koncept vyššie v rámci úplne usporiadaného konceptového zväzu.

FrstUpprCncpt - pomocou premennej session a metódy getAttribute boli získané vyhľadané koncepty a používateľov dopyt. Výstupom zobrazeným na obrazovke výstupu boli dokumenty, ktoré sa vyskytovali v extente konceptu, ktorý sa v konceptovom zväze nachádzal ako prvý o úroveň vyššie od najpodobnejšieho konceptu voči dopytu. Ak používateľ nebol spokojný s daným výstupom, mohol do formulára zadať nové kľúčové slova alebo mohol ísť o úroveň vyššie, čiže mohol zobraziť neredundantné dokumenty vyskytujúce sa v konceptoch, ktoré sa nachádzajú o úroveň vyššie od najpodobnejšieho konceptu k dopytu.

FrstLwrCncpt - pomocou premennej session a metódy getAttribute boli získané vyhľadané koncepty a používateľov dopyt. Výstupom zobrazeným na obrazovke výstupu bol dokument, resp. dokumenty, ktoré sa vyskytovali v extente konceptu, ktorý sa v konceptovom zväze nachádzal ako prvý o úroveň nižšie od najpodobnejšieho konceptu voči dopytu. Ak používateľ nebol spokojný s daným výstupom, mohol do formulára zadať nové kľúčové slova.

UpprCnts – pomocou premennej session a metódy getAttribute boli získané vyhľadané koncepty a používateľov dopyt. Výstupom zobrazeným na obrazovke výstupu boli dokumenty, ktoré sa vyskytovali v extentoch konceptov, ktoré sa v konceptovom zväze nachádzali úroveň vyššie od najpodobnejšieho konceptu voči dopytu. Ak používateľ nebol spokojný s daným výstupom, mohol do formulára zadať nové kľúčové slova.

Pre otestovanie bolo uskutočnených niekoľkých experimentov použitím vytvorenej aplikácie pre vyhľadávanie webových stránok pomocou formálne konceptuálnej analýzy. V každom experimente bol použitý iný prístup pre vyhľadanie dokumentov k danému dopytu. Efektivita vyhľadávania bola kvantifikovaná pomocou funkcie presnosti a návratnosti. Pre výpočet týchto dvoch ukazovateľov efektivity vyhľadávania bolo nutné si dopredu určiť, ktoré dokumenty z danej množiny dokumentov sú relevantné k dopytu. Na základe vyhľadaných výsledkov bolo možné následne kvantifikovať presnosť a návratnosť pre daný dopyt.

Z vykonaných experimentov bolo možné vyvodiť, že vyhľadanie najpodobnejšieho konceptu k dopytu predstavuje najlepší kompromis medzi presnosťou a návratnosťou. Ak sa však chceme pokúsiť zvýšiť návratnosť, tak môžeme pristúpiť ku nejakému konceptu (v našom prípade k prvému), ktorý sa nachádza v hierarchii vyššie od najpodobnejšieho konceptu. Ak chceme maximalizovať návratnosť, tak môžeme vyhľadať všetky koncepty, ktoré sa nachádzajú o úroveň vyššie v hierarchii alebo všetky koncepty, ktoré boli vyhľadané ako najpodobnejšie k dopytu. Pri týchto postupoch je vysoko pravdepodobné, že sa návratnosť zvýši, ale daňou za to býva výrazný pokles presnosti. Pri bázach dát s vysokým počtom dokumentov môžeme dostať na výstup veľké množstvo dokumentov. V tomto prípade je vhodné na výstup zobraziť nejaký z konceptov (v našom prípade prvý), ktorý sa nachádza v hierarchii konceptového zväzu nižšie. Zobraziť všetky koncepty, ktoré sa nachádzajú v hierarchii nižšie nemá zmysel, keďže by sme znova dostali koncept, z ktorého sme vychádzali. Zmysel by malo napr. iba zlúčenie niekoľkých konceptov, ktoré sa nachádzajú nižšie. Pri vyhľadávaní konceptov, ktoré sa nachádzajú od najpodobnejšieho konceptu nižšie výrazne klesá návratnosť. Presnosť v takomto prípade môže vzrásť, ale môže sa stať (ako sa stalo aj v našom prípade), že práve relevantný dokument sa týmto prístupom nezobrazí a tým pádom môže presnosť byť aj nulová.

POĎAKOVANIE

Táto práca bola podporená realizáciou projektu Rozvoj Centra informačných a komunikačných technológií pre znalostné systémy (ITMS kód: 26220120030) na základe podpory operačného programu Výskum a vývoj financovaného z Európskeho fondu regionálneho rozvoja.

LITERATÚRA

[1] R. Baeza-Yates, B. Ribeiro-Neto. *Modern information retrieval*. Harlow, Addison Wesley, 1999.
[2] P. Bednár, P. Butka, J. Paralič, "Java library for support of text mining and retrieval," In: Znalosti 2005, Sborník příspěvků 4. ročníku konference, 2005, pp.162-169.
[3] P. Butka, J. Pocs, J. Pocsova, "Design and implementation of incremental algorithm for creation of generalized one-sided concept lattices," In: CINTI 2011, 12th IEEE International Symposium on Computational Intelligence and Informatics, IEEE conference, Budapest, Hungary, 2011, pp. 373-378.
[4] B. Ganter, R. Wille. *Formal Concept Analysis*. Springer, Berlin, 1999.
[5] T. Kuzar, P. Navrat, "Preprocessing of Slovak Blog Articles for Clustering," In: International Conference on Web Intelligence and Intelligent Agent Technology, Toronto, IEEE, 2010, pp. 314-317.
[6] J. Paralič a kol. *Dolovanie znalostí z textov*. Equilibria, Košice, 2010.

Vyhľadávanie webových stránok pomocou ekvivalencie tried v oblasti rough sets

[1]*Martin SARNOVSKÝ,* [1]*Maroš MARCINOV,* [2]*Peter BUTKA*

[1] Katedra kybernetiky a umelej inteligencie, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika
[2] Centrum pre informačné technológie, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

[1] martin.sarnovsky@tuke.sk, [1] maros.marcinov@student.tuke.sk, [2] peter.butka@tuke.sk

***Abstrakt* — Táto práca sa zaoberá problematikou vyhľadávania webových stránok použitím prístupu ekvivalencie tried z oblasti rough sets. Podáva teoretický pohľad na problematiku vyhľadávania informácií a prístupu rough sets, detailnejšie popisuje princíp ekvivalencie tried. Cieľom tejto práce implementácia aplikácie založenej na prístupe ekvivalencie tried a jej využitie v oblasti vyhľadávania informácií a jej otestovanie.**

***Kľúčové slová* — ekvivalencia tried, rough sets, vyhľadávanie, spracovanie textov**

I. Úvod

Vyhľadávanie informácií (z angl. Information Retrieval, IR) môžeme definovať ako počítačom riadený výber skupiny dokumentov z celého korpusu dokumentov, ktoré sú používateľovi vrátené ako odpoveď na jeho požiadavku [2]. Požiadavka alebo dopyt je minimálne jedno kľúčové slovo podľa ktorého sa vykonáva vyhľadávanie v dokumentoch. Na vyhľadávanie informácií preto môžeme pozerať ako na klasifikátor, ktorý rozdeľuje dokumenty z korpusu na dve skupiny. Jednu skupinu tvoria dokumenty, ktoré sú používateľovi zobrazené preto že sú pre neho relevantné. Druhú skupinu tvoria dokumenty v ktorých je nulová zhoda s dopytom, tým pádom sú pre užívateľa irelevantné a nie sú ani zobrazené. Na systémy ktoré neposkytujú len skupinu dokumentov zaujímavých pre užívateľa, ale poskytujú aj ohodnotenie pre každý dokument, sa potom možno pozerať ako na systémy, ktoré počítajú stupeň príslušnosti dokumentu do triedy relevantných dokumentov. Z tohto pohľadu je zaužívaný pojem vyhľadávanie dokumentov. Užívateľ často zadáva dopyt vo forme boolovského výrazu, ako napr.: (slnečno AND teplo) OR dážď. Snahou najnovších výskumov je aby užívateľ mohol zadávať tzv. užívateľsky priateľský dopyt, tedy v prirodzenom jazyku. Proces vyhľadávanie informácií je zložený zo siedmych krokov, ktoré sú graficky znázornené na obrázku.

Vyhľadávanie informácií sa dá definovať nasledujúcou štvoricou: [D, Q, F, R(qj, dj)]. V tomto výraze D vyjadruje množinu logických pohľadov, t.j. rôzne reprezentácie dokumentov z korpusu, Q je množina logických pohľadov informačných požiadaviek čiže dotazov, F je prostredie pre modelovanie reprezentácií dokumentov, dopytu a vzťahom medzi nimi, R(qj, dj) vyjadruje funkciu ohodnotenia, ktorá priradí reálne číslo daného dotazu a dokumentu . Funkcia je definíciou usporiadania medzi dokumentmi vzhľadom na daný dopyt.

Vyhodnotenie systémov pre vyhľadávanie informácií môžeme robiť podľa rôznych faktorov, napr. kvalita vyhľadávania, časové hľadisko alebo jeho efektívnosť. Časové hľadisko je pomerne jednoducho merateľné a nie je zložité ani jeho vyhodnotenie. Ale kvalita vyhľadávania je už oveľa viac komplexnejšie hodnotenie.

Dve základné miery kvality sú návratnosť (recall) a presnosť (precision). Nech R je množina, ktorá obsahuje všetky relevantné dokumenty k dotazu I a nech A je množina dokumentov vrátených systémom [5]. Ra označuje prienik medzi R a A.

Návratnosť je definovaná ako podiel nájdených relevantných dokumentov ku všetkým relevantným:

$$Návratnosť = \frac{|R_a|}{|R|}$$

Presnosť je podiel nájdených relevantných dokumentov ku všetkým, ktoré systém našiel:

$$Presnosť = \frac{|R_a|}{|A|}$$

Zvyčajne sa kvalita reprezentuje grafom závislosti presnosti od návratnosti, pričom sa aproximujú hodnoty presnosti pre 11 hodnôt návratnosti, a to 0%, 10% ... 90%, 100%.

Ak chceme vyhodnotiť kvalitu vyhľadávania v jednom čísle tak použijeme tzv. R-presnosť. R-presnosť je presnosť na R-tej pozícií návratnosti, pričom R je počet dokumentov skutočne relevatných k dotazu.

II. EKVIVALENCIA TRIED V OBLASTI ROUGH SETS

Rough sets boli predstavené v skorých 90-tych rokoch dvadsiateho storočia Zdislawom I. Pavlakom. Teória rough sets hovorí, že množinu objektov, o ktorých máme nekompletné informácie (teda nemáme k dispozícií všetky atribúty, ktoré popisujú objekt) vieme začleniť do ekvivalentných tried [1]. To znamená, že v jednej ekvivalentnej triede sa budú nachádzať objekty navzájom od seba neodlíšiteľné. Môžeme teda povedať že neurčitosť je nevyhnutnou vlastnosťou, ktorá je podmienkou pre aplikovanie rough sets. Veľmi vhodná vyzerá byť aplikácia rough sets do oblasti medicíny, kde symptómy reprezentujú atribúty zdravotného stavu pacienta, na základe ktorých sa vykoná rozhodovanie. Je vhodná aj v informatike pri klasifikácií a analýze (nekompletných) dát.

Matematicky vychádzame z aproximačného priestoru, ktorý reprezentuje dvojica $A = (U, R)$ [3], kde:

U – množina nazývaná univerzum,

R – binárna relácia na R, označovaná ako relácia nerozlíšiteľnosti.

Pre ďalšie matematické vyjadrenia potrebujeme poznať tieto pojmy:

X – cieľová podmnožina prvokov, $X \subseteq U$,

x – prvok alebo objekt, $x \subseteq U$,

$[x]R$ - R-elementárna množina obsahujúca prvok x.

Metóda rough sets rozdelí univerzum na dri základné podmnožiny a to: dolná aproximácia, horná aproximácia a hraničný región [4].

Dolná aproximácia je podmnožina prvkov o ktorých vieme s istotou povedať že patria do cieľovej podmnožiny X.

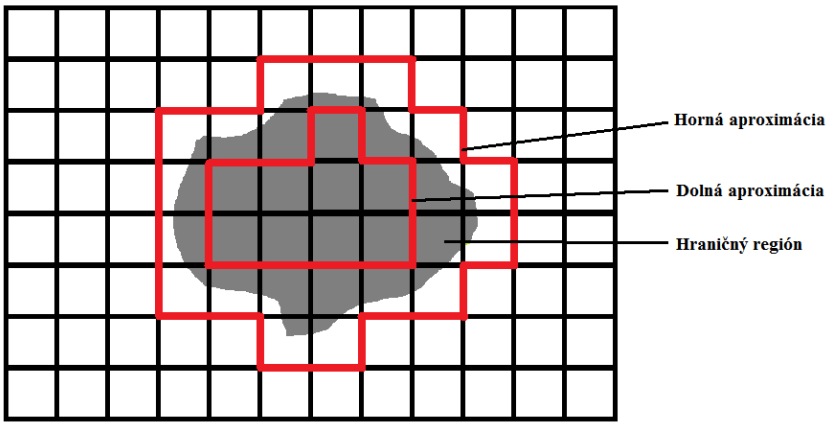
$$\underline{R}X = \{x \in U; [x]R \in X\}$$

Horná aproximácia je podmnožina prvkov o ktorej vieme s istotou povedať že neobsahuje prvky, ktoré nepatria do cieľovej podmnožiny X.

$$\overline{R}X = \{x \in U; [x]R \cap X \neq 0\}$$

Hraničný región je podmnožina, ktorá vznikne odčítaním dolnej aproximácie od hornej.

$$BN_R(X) = \overline{R}X - \underline{R}X$$



Obr. 1 Znázornenie 3 základných podmnožín podľa rough sets

III. VYHĽADÁVANIE S POUŽITÍM MODELU NA BÁZE EKVIVALENTNÝCH TRIED

Keďže hlavným cieľom bol návrh, implementácia a otestovanie IR (Information Retrieval) systému pre vyhľadávanie informácií z webových stránok použitím prístupu formálnej konceptuálnej analýzy, model vytvorený podľa postupu prezentovaného v predchádzajúcej kapitole bol použitý pre vyhľadávanie. Na predspracovanie, indexáciu vstupných dát a tvorbu modelu bola využitá javovská softvérová knižnica JBowl (Java Bag-of-words Library), ktorá je bližšie popísaná v [2]. Pre potreby zostavenia požadovaného modelu bol z projektu GOSCL [3] získaný inkrementálny algoritmus pre tvorbu zovšeobecnených jednostranných konceptových zväzov a implementovaný do knižnice JBowl tak, aby využíval jeho štruktúry. Webová aplikácia na vyhľadávanie webových stránok využíva technológiu JSP (Java Server Pages).

Databáza je vo formáte XML tvorí ju 50 kníh. Presnejšie 50 recenzií kníh, ktoré sme získali zo stránky http://www.goodreads.com. Koreňový element <dataset> obsahuje 50 <document> elementov s atribútmi name a id. Každý element <document> má minimálne 6 ďalších podelementov. A to <title>, <author>, <category>, <date>, <rating> a <text>.

Jednotlivé elementy obsahujú nasledujúce informácie:

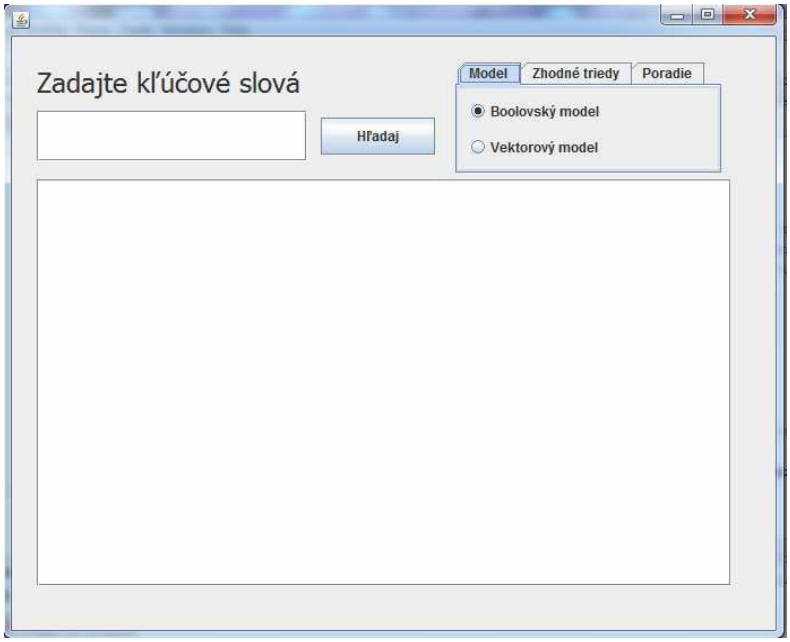
- <title> - titul teda názov diela,
- <author> - autor, ktorý knihu napísal,
- <category> - literárny druh alebo kategória, do ktorej dielo patrí, keďže kniha sa dá
- zaradiť do viacerých kategórií, tak aj element document môže obsahovať podľa potreby
- viac elementov category,
- <date> - dátum, kedy bola recenzia zverejnená na stránke http://www.goodreads.com,
- <rating> - je hodnotenie knihy čitateľmi a je v rozsahu jedna až päť, pričom päť je
- najvyššie hodnotenie,
- <text> - obsahuje samotný text recenzie.

Navrhnutý IR systém je implementovaný použitím knižnice JBowl. Po predspracovaní databázy sa vytvorí dokument-term matica. V tomto modeli sa váham s nulovými hodnotami priradí nula a váham s nenulovými hodnotami sa priradí číslo 1, v prípade boolovského modelu, alebo čísla 1 až 3 v závislosti na hodnote váhy termu. Používateľ bude mať možnosť voľby medzi jedným z týchto modelov. Každý dokument bude teda reprezentovaný vektorom s hodnotami 0 až 1, prípadne 0 až 3.

Hodnoty jednotlivých atribútov vektora vlastne budú zodpovedať relevantnosti dokumentov vzhľadom k dopytu. Čím sú hodnoty vyššie, tým je term vzhľadom na dokument relevantnejší. Je veľmi pravdepodobné, že sa v databáze bude vyskytovať viac dokumentov so zhodnými vektormi. Takého dokumenty, ktoré budú od seba navzájom neodlíšiteľné, potom môžeme zaradiť do jednej ekvivalentnej triedy. Takto vytvorené ekvivalentné triedy následne zoradíme od najrelevantnejších po menej relevantné podľa súčtu hodnôt vektora. Avšak môže nastať prípad, že hodnoty niektorých tried budú rovnaké (napr. pri dopyte, ktorý sa skladá z troch kľúčových slov môžu vzniknúť ekvivalentné triedy, ktorých dokumenty budú reprezentovať

nasledovné vektory: [1 1 1],[1 1 0],[0 1 1],[1 0 1]. Hodnoty tried 2, 3 a 4 sú zhodné). V takomto prípade bude mať používateľ na výber. Buď takto zhodné triedy spojí do jednej väčšej alebo ich zoradí podľa priemerného ratingu, ktorý sa vypočíta z ratingov dokumentov v príslušných triedach.

Keď budú ekvivalentné triedy zoradené podľa relevantnosti, tak už bude ostávať iba zoradiť dokumenty v jednotlivých triedach. Opäť bude mať používateľ na výber z dvoch možností. Dokumenty sa zoradia buď podľa ratingu uvedeného pre každý dokument v databáze. V tomto prípade budú uprednostnené dokumenty s lepším ratingom (hodnotením čitateľov) pred dokumentmi, ktoré majú síce zhodný výskyt kľúčových slov, ale čitatelia im pridelili nižšie hodnotenie. Druhá možnosť je, že sa dokumenty zoradia podľa súčtu pôvodných váh z matice dokument-term. Pri tejto možnosti sa berie citlivejšie do úvahy výskyt kľúčových slov v dokumentoch. K aplikácií bolo vytvorené GUI (Graphical User Interface) prostredníctvom ktorého používateľ zadáva kľúčové slová a volí nastavenia vyhľadávania spomenuté vyššie. Aplikácia zobrazuje všetky vrátené dokumenty okrem tých, v ktorých sa nevyskytuje ani jedno z kľúčových slov.

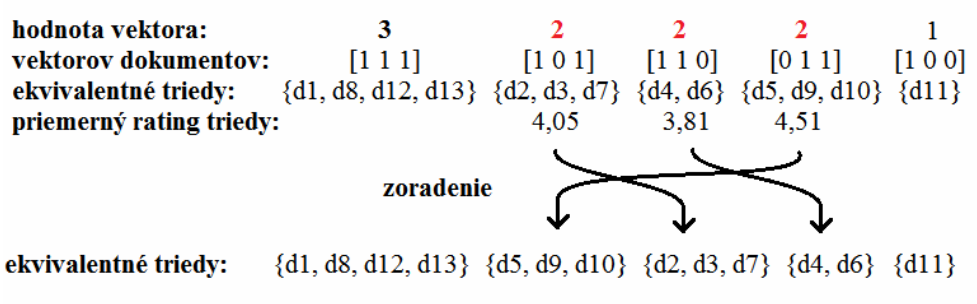


Obr. 2 Grafická rozhranie aplikácie

Ako je uvedené v návrhu aplikácie, tak najprv potrebujeme získať maticu termdokument. Na tento cieľ bola vytvorená trieda InstanceValues. Nachádzajú sa v nej metódy, ktoré boli naprogramované v pôvodnom projekte JBowl. Zadaný dopyt od užívateľa v podobe kľúčových slov sa porovná so zoznamom zaindexovaných termov, ktorý sa vytvoril počas fáz predspracovania v metóde comparem, ktorá je implementovaná v triede KeyWords. Metóda vráti zoznam indexov termov, ktoré sa zhodujú s kľúčovými slovami. Podľa tohto zoznamu indexov sa z matice termdokument vyberú už iba tie stĺpce, ktoré zodpovedajú kľúčovým slovám zadaným od užívateľa. Následne sa vytvorí boolovský alebo vektorový model. To, o aký model pôjde, si užívateľ navolí v záložke Model, ktorá sa nachádza v zozname záložiek v pravej hornej časti grafického užívateľského rozhrania. Okno grafického rozhrania môžeme vidieť na obrázku Obr. 2. Model reprezentácie obsahuje vektory dokumentov, ktoré znázorňujú prítomnosť alebo neprítomnosť kľúčových slov. Dokumenty reprezentované vektormi môžu byť od seba navzájom nerozoznateľné, t. j. ekvivalentné. Všetky dokumenty sa rozdelia do ekvivalentných tried. V jednej triede sa bude nachádzať podmnožina dokumentov, ktoré sú zhodné. Vytvorenie ekvivalentných tried sa v aplikácií vykoná v triede EqualentClasses. Konkrétne v metóde makeEqualentClasses. Metóda je bez návratovej hodnoty, výsledné triedy sa vložia do triednej premennej equalentClasses. Metóda má jeden vstupný parameter, ktorým je už spomínaný boolovský alebo vektorový model.

Vytvorené ekvivalentné triedy si zoradíme tak, aby boli užívateľovi poskytnuté najprv najrelevantnejšie dokumenty. Poradie si určíme podľa hodnoty, ktorú dostaneme spočítaním prvkov vektora, ktorý reprezentuje dokumenty istej triedy. Ak máme napríklad ekvivalentnú triedu, ktorej boolovský vektor má tvar [1 1 0 1], tak táto trieda bude užívateľovi vypísaná skôr

ako trieda, ktorej dokumenty majú vektor [1 0 0 1]. Pretože sa v dokumentoch tejto triedy nachádza viac kľúčových slov, zoradenie tried sa robí metódou buble sort. Na zoraďovanie ekvivalentných tried podľa relevantnosti je v triede EqualentClasses implementovaná metóda sort. Zoraďovaním tried podľa hodnôt vypočítaných zo súčtu prvkov vektora môže nastať prípad, že budeme mať viac tried s rovnakou hodnotou. Pre takýto prípad si užívateľ ešte pred vyhľadávaním nastaví v záložke Zhodné triedy. Radio button Spojiť spojí triedy s rovnakou hodnotou do jednej väčšej triedy. V prípade, že si užívateľ zvolí Zoradiť podľa ratingu, tak sa vypočíta priemerný rating pre každú zo zhodných ekvivalentný tried a poukladajú sa podľa ratingu. Tento spôsob je znázornený na obrázku Obr. 3.



Obr. 3 Zoradenie zhodných tried podľa priemerného ratingu

Keď máme zoradené ekvivalentné triedy tak pred tým ako sa zobrazia užívateľovi, musíme zoradiť dokumenty v každej triede. Pre prácu s dokumentmi v rámci ekvivalentnej triedy je v aplikácií naprogramovaná trieda Documents. Pri vytváraní inštancie triedy sa v konštruktore triedy nastavia premenné triedy, aby sa mohli použiť metódy implementované v triede Documents. Zoraďovanie dokumentov v rámci každej ekvivalentnej triedy vykonáva metóda sortDocuments. Zoraďovanie sa môže vykonať dvoma spôsobmi. Užívateľ si v okne grafického rozhrania (Obr. 7) v záložke poradie môže vybrať radio button Váha termu. V tomto prípade sa pristupuje k váham z pôvodnej matice term-dokument, podľa ktorých sa poukladajú dokumenty v rámci každej triedy. Pri tomto spôsobe sa berie citlivejší ohľad na výskyt kľúčových slov v dokumentoch. Druhý radio button je Rating. Pri zadaní tejto voľby sa vykoná metóda, ktorá pomocou parsera zistí rating pre každý dokument a dokumenty v rámci triedy zoradí od najväčšieho po najmenší. Výhodou tohto spôsobu je, že sa užívateľovi skôr odporúčajú knihy s lepším hodnotením od čitateľov ako tie knihy, ktoré majú v modelovej reprezentácií rovnaké zastúpenie kľúčových slov, ale majú horšie hodnotenie od čitateľov. Takto zoradené dokumenty sa užívateľovi zobrazia v GUI do okna pre výstup (Obr. 2) ako zoznam kníh. Postupne sa zobrazí názov diela, meno autora a úryvok z textu. Do výstupného okna sa nezobrazuje trieda, ktorej dokumenty majú vektory s nulovými hodnotami, pretože neobsahujú žiadne kľúčové slovo.

Na vyhodnotenie efektívnosti IR systému sme použili typové dotazy, ktorými sme testovali vyhľadávanie pri všetkých kombináciách nastavenia vyhľadávania. Najlepšie poradie dokumentov je dosiahnuté pri nastaveniach Model: Vektorový, Zhodné triedy: Spojiť a Poradie: podľa Ratingu. Je to optimálne nastavenie, ktoré dáva užívateľovi výsledok, ktorý je najlepším kompromisom medzi kvalitou odporúčaných kníh (meraná ratingom) a oblasťou o ktorú má záujem (podľa kľúčových slov, ktoré zadal). Najhoršie výsledky boli dosiahnuté pri nastaveniach Model: Boolovský, Zhodné triedy: Zoradiť podľa ratingu a Poradie: podľa Váh termov.

POĎAKOVANIE

Táto práca bola podporená realizáciou projektu Rozvoj Centra informačných a komunikačných technológií pre znalostné systémy (ITMS kód: 26220120030) na základe podpory operačného programu Výskum a vývoj financovaného z Európskeho fondu regionálneho rozvoja.

LITERATÚRA

[1] POLKOWSKI, Lech: Rough Sets: Mathematical Foundations. Advances in Soft Computing, Physica–Verlag, Heidelberg, 2002. 554 s. ISBN 3-7908-1510-1.

[2] PARALIČ, Ján, et al. 2010. Dolovanie znalostí z textov. Košice: Equilibria, s.r.o., Košice, 2010. 183 s. ISBN 978-80-89284-62-7.

[3] DVOŽAK, Vladimír: Modelování algoritmu – generování rozhodovacích pravidel pomocí rough množín. Diplomová práca. Pardubite: UPa, FSE, 2009. 61 s.B. Ganter, R. Wille. *Formal Concept Analysis*. Springer, Berlin, 1999.

[4] NOVOTNÝ, Jiři: Stroje a množiny Zdzdislawa Pawlaka [online]. Brno: VUT, FAST, 2006. [s.a.]. [cit. 2012-4-2]. Dostupné na internete: <http://math.fce.vutbr.cz/~pribyl/workshop_2006/prispevky/Novotny.pdf>.

[5] KOŠTIAĽ, Ivan: Spracovanie a kategorizácia textových dokumentov pre účely semiautomatického linkovania na znalostný model: Diplomová práca. Košice: TU, FEI, 2002. 55.

Jednoduchý príklad modelovania autonómneho agenta v simulačnom prostredí Netlogo

[1]*Martin SARNOVSKÝ,* [1]*Patrik VINCEJ,* [2]*Peter BUTKA*

[1]Katedra kybernetiky a umelej inteligencie, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika
[2]Centrum pre informačné technológie, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

[1]martin.sarnovsky@tuke.sk, [1]patrik.vincej@student.tuke.sk, [2]peter.butka@tuke.sk

***Abstrakt* — Táto práca popisuje rôzne typy agentov, ich vlastnosti a správanie sa vo svojom prostredí, ako aj konkrétne príklady agentov simulovaných v rámci prostredia Netlogo pre pohyb v priestore bez a s prekážkami. Správanie sa rôznych typov agentov závisí od ich vlastností, niekoľko príkladov bude v tomto článku predstavených. Cieľom bolo modelovať rôzne typy jednoduchých agentov pre prejdenie celého priestoru prostredia.**

***Kľúčové slová* — agent, agentové simulácie, Netlogo, simulačné prostredie**

I. INTRODUCTION

V praktických aplikáciách existuje mnoho príkladov systémov prvkov, ktoré vykazujú komplexitu a emergentné správanie na základe pomerne jednoduchých rozhodovacích pravidiel. Tieto sa často nedarí odhaliť štandardným analytickým postupom, ale používa sa postup založený na simulácii. Logickým prostriedkom takéhoto skúmania je prepojenie modelovania a simulácie s agentovými technológiami. Cieľom agentových simulácií je potom popísať jednotlivé prvky systému a ich správanie v reči agentov, ich spustenie v simulačnom prostredí, ako aj analýza výsledkov a priebehu simulácií [1]. Praktické skúsenosti a znalosti ohľadom typov agentov a ich správania v simulačnom prostredí preto poskytujú elementárnu znalosť, ktorá pomáha pri modelovaní a praktickej aplikácii v zložitejších prípadoch.

Cieľom tohto článku je popísať základné pojmy oblasti agentových simulácií (v kapitole II), simulačné prostredia a najmä konkrétne prostredie Netlogo (v kapitole III), ako aj príklad jednoduchého (autonómneho) agenta (respektíve viacerých typov) pre pohyb v priestore a jeho úplne prejdenie, pričom toto prostredie môže obsahovať prekážky. Odpovedá to klasickej aplikácii známej aj v praxi ako autonómny vysávač (kapitola IV). V závere sú spomenuté niektoré možností na rozšírenia, ako aj zoznam relevantnej literatúry.

II. AGENTOVÉ SYSTÉMY

Pojem agentový systém sa uvádza prevažne v spojení s prostredím, v ktorom pôsobí agent, čiže aktívny prvok, ktorý je vybavený určitou dávkou inteligencie. Tá mu umožňuje riešiť zadaný problém [7]. Najčastejšie používaná definícia agenta (Wooldridge&Jennings) rozlišuje dva typy - slabý a silný [5].

Slabý agent predstavuje hardvérový alebo softvérový počítačový systém, ktorý spĺňa nasledujúce požiadavky. Prvou je autonómnosť, kde takýto agent koná bez akéhokoľvek priameho ľudského alebo iného zásahu a má kontrolu nad svojimi akciami a vnútorným stavom. Druhou je sociálna schopnosť - agent komunikuje s inými agentmi, alebo s človekom prostredníctvom nejakého komunikačného jazyka. Ďalšou je reaktivita, kde agent vníma svoje okolie a vo vhodnom čase reaguje na zmenu vo svojom okolí. Poslednou je proaktivita, kde pri konaní agentov nie je odozva jednoduchá na stav okolia, ale môže byť aj cieľovo orientovaná s prevzatím iniciatívy. Silný agent má navyše k predchádzajúcim vlastnostiam aj schopnosti spracovať ďalšie pojmy, ktoré nám vystihujú mentálne a emocionálne stavy.

Rôzne ďalšie atribúty sú diskutované v rámci agentov. Prvou je mobilita, čiže je to schopnosť agenta pohybovať sa v elektronickej sieti. Druhou je pravdivosť (pravdovravnosť),

je tu predpoklad, že agenti nebudú vedome oznamovať nesprávne informácie. Treťou je zhovievavosť, tu je predpoklad, že agenti nemajú konfliktné ciele a každý agent sa bude vždy snažiť robiť to o čo bude požiadaný. Poslednou je rozumnosť, čiže tu sa predpokladá, že agent bude konať v záujme dosiahnutia svojich cieľov a nebude konať do takej miery, aby došlo k cieľu, ktorý už bol dosiahnutý.

Táto definícia od autorov Wooldridge a Jennings bola prepracovaná v práce Nwanu [3], ktorý definoval pojem agent prostredníctvom troch vlastností. Takýto agent musí spĺňať aspoň dve z týchto troch vlastností:

- Autonómnosť - agent má schopnosť pohybovať sa samostatne v prostredí bez prispenia človeka aj vtedy ak jeho okolie nie je popísané, kde dôležitou časťou autonómnosti je proaktivita, čiže konanie agenta je cieľovo orientované s prevzatím iniciatívy.
- Kooperácia - agent komunikuje s inými agentmi alebo s človekom prostredníctvom nejakého komunikačného jazyka, táto vlastnosť patrí medzi dôležité vlastnosti multiagentových systémov a podľa [4] sú agenti schopný vykonávať svoje akcie aj bez vzájomného pôsobenia.
- Učenie - aby agent mohol reagovať a konať vo svojom prostredí, ktoré má byť dynamické a nedeterministické, musí sa zaviesť faktor učenia cez vzájomné pôsobenie agenta a prostredie. Existuje predpoklad, že učenie môže postupom času zlepšiť kvalitu v správaní agenta v prostredí.

Podľa atribútov (vlastností) môžme určiť typy agentov. Asi najčastejšie uvádzané rozdelenie agentov je podľa Nwanu [3]. *Spolupracujúci agenti* sú charakterizovaní hlavne tým, že majú schopnosť vzájomne komunikovať s tým cieľom, že spolu ako celok presahujú schopnosti agentov jednotlivcov. Mali by sa racionálne správať a mali by byť schopní samostatne existovať v otvorenom multiagentovom systéme. *Agenti rozhrania* boli vytvorení ako osobní asistenti, ktorí spolupracujú s používateľom v tom istom prostredí. Od spolupracujúcich agentov sa odlišujú hlavne tým, že pri komunikácii s používateľom nepotrebujú špeciálny komunikačný jazyk. Používateľovi títo agenti pomáhajú pri učení ovládať určitú aplikáciu. Takýto agent dokáže dohliadať na akcie vykonania úlohy. *Mobilní agenti* sú charakteristickými atribútmi autonómnosť a mobilita. Po svojom vzniku majú schopnosť samostatne rozhodovať, ktoré miesta preskúmajú vo svojom prostredí a aké akcie v ňom vykonajú. *Informační (internetoví) agenti* dokážu pomôcť spracovať narastajúci objem informácií, ktorý zachytávame teraz a v budúcnosti bude narastať. Títo agenti pracujú v úlohách riadenia, spracovania a porovnávania informácií z mnohých distribuovaných zdrojov. *Reaktívny agent* vznikol na základe slabej výkonnosti systémov používajúcich náročné symbolické reprezentácie vlastného stavu a stavu okolia, nad ktorými sú vykonávané procesy, ktoré vedú k dosiahnutiu požadovaného správania agenta. Dokážu vnímať svoje okolie a na základe stavu okolia reagujú. *Hybridní agenti* kombinujú agentov dvoch alebo viacerých typov jednej skupiny agentov do jedného agenta. *Heterogénny agent* integruje minimálne dva agentové systémy, ktoré patria do rozdielnych skupín agentov. *Rozumní agenti* predstavujú plnú kombináciu troch základných vlastností agentov: autonómnosť, kooperácia a učenie. *Deliberatívny agent* je typ agenta, ktorý je schopný uvažovať o svojich dlhodobých zámeroch. Môžeme ho nazvať aj uvažujúcim agentom. Tento agent si dokáže uchovať symbolickú reprezentáciu prostredia, vnútorných stavov a znalostí. Na ich základe vytvára plány pre dosahovanie cieľov. *Sociálny agent* udržuje a rozširuje poznatky o predošlých agentoch a plánoch. Je schopný uvažovať o ich cieľoch a motiváciách. Tieto informácie majú po väčšine charakter adries, názvov agentov, špecifikácie ich schopností, čo im umožňuje spoluprácu vzájomných aktivít. Ďalej agent môže mať okrem toho často k dispozícii aj históriu predchádzajúcich vzájomných pôsobení, ktoré mu v budúcich akciách pomôžu lepšie vyhľadať pomoc ostatných, poprípade zmierniť negatívne dôsledky obmedzenosti jeho vlastných zdrojov.

Ďalším dôležitým aspektom je *prostredie*, ktoré predstavuje všetko to, s čím agent prichádza do styku počas svojej činnosti [5]. Prostredie z hľadiska agenta môžeme rozdeliť na plne pozorovateľné (agent môže senzormi sledovať jeho kompletný stav) alebo čiastočne pozorovateľné (agent nie je schopný sledovať všetky stavy prostredia). Prostredie je pre agenta statické, ak sa môže meniť iba jeho akciami. Prostredie je pre agenta dynamické, ak sa môže meniť aj bez zásahu agenta. Prostredie je deterministické ak je jeho stav po vykonaní nejakej akcie daný iba touto akciou a predchádzajúcim stavom tohto prostredia. Prostredie je diskrétne vtedy ak má konečne alebo spočetne mnoho stavov. Pre agenta založeného na číslicových počítačoch je každé prostredie diskrétne. Prostredie je spojité, ak má nekonečný počet stavov.

III. SIMULAČNÉ PROSTREDIE - NETLOGO

V multiagentových systémoch je zoskupená množina agentov za určitým cieľom, medzi ktorými sú definované konkrétne vzťahy a väzby. Každý agent má určité prostriedky, s ktorými môže disponovať a zároveň aj obmedzenia, v hraniciach ktorých sa musí pohybovať pri dosiahnutí cieľa. Každý z agentov sa snaží o čo najkvalitnejšie uskutočnenie danej úlohy, čo sa v konečnom dôsledku prejaví v celkovom riešení. V takejto "spoločnosti agentov", agenti spoločne koordinujú svoje schopnosti, znalosti, ciele a plány za účelom vykonania nejakej akcie alebo kolektívneho vyriešenia nejakého problému. Títo agenti môžu pracovať buď spoločne na jednej úlohe alebo každý samostatne na inej alebo rovnakej úlohe. Multiagentová simulácia spočíva vo využití tzv. agentov, ktorí predstavujú autonómne entity s relatívne jednoduchým chovaním, reprezentujúce reálne jednotky sledovaného systému, situované do definovaného kontextu prostredia, v ktorom jednajú a reagujú.

Praktická realizácia simulácií zahŕňa proces modelovania a nástroje používané na modelovanie a simuláciu. Ako nástroje používané pre tieto potreby uveďme niekoľko príkladov. Jednoduchý, praktický a názorný (vhodný najmä pre vzdelávacie účely, s dobrou a obšírnou dokumentáciou jazyka pre popis agentov a existujúcich modelov) je Netlogo (http://ccl.northwestern.edu/netlogo/). Rýchly a populárny nástroj pre tvorbu multiagentových systémov s voliteľnou sadu vizualizačných nástrojov v oblasti 2D a 3D predstavuj MASON (http://cs.gmu.edu/~eclab/projects/mason/). Výborné prostredie pre tvorbu modelov vo viacerých jazykoch (C++, Java, Groovy) so silným vývojovým prostredím poskytuje Repast (existuje v rôznych verziách, http://repast.sourceforge.net/). Voľne dostupné Java prostredie pre vývoj multiagentových systémov vo všeobecnejšom ponímaní pokytuje JADE (http://jade.tilab.com/). Mnohé z týchto nástrojov poskytujú integrované prostriedky pre rýchly návrh a vývoj simulačných modelov, ako aj ich samotné spúšťanie a analýzu dosiahnutých výsledkov.

Netlogo je multiagentové simulačné prostredie napísané v jazyku Java, ktoré vzniklo v roku 1999 a jeho autorom je Uri Wilensky. Netlogo sa hodí na modelovanie systémov, ktoré sa vyvíjajú priebežne v čase a tak môže užívateľ zadávať príkazy a sledovať správanie sa jednotlivých agentov na mikro a makro úrovni. Netlogo umožňuje užívateľom modifikovať už existujúce modely a skúmať správanie sa agentov v rôznych podmienkach. Umožňuje takisto vytvárať nové vlastné modely. K Netlogu existuje veľmi dobrá a rozsiahla dokumentácia s návodmi a už aj hotovými modelmi, ktoré je možné používať a pozmeňovať. Netlogo je možné pripojiť aj k výukovému nástroju Hubnet. Netlogo vychádza z nástroja Starlogo oproti ktorému má nové vlastnosti, funkcie a lepšie jazykové a užívateľské rozhranie.

Netlogo beží na rôznych platformách, či už je to Mac, Linux, alebo Windows. Jazyk Netloga je plne programovateľný, má jednoduchú jazykovú štruktúru. Pohyblivý agenti (korytnačky) sa pohybujú po mriežke nehybných agentov (políčka) a je možné ich prepojiť a tak z pohyblivých agentov môžu vzniknúť zhluky, siete, alebo grafy. Modely sa dajú reprodukovať na rôznych platformách. Prostredie môže reprezentovať modely v 2D aj 3D zobrazení, umožňuje meniť veľkosť a otáčať vektorové tvary. Pri jednotlivých agentoch sú dané popisy agentov. Počas behu simulácie je dostupný príkazový panel určený na komunikáciu v reálnom čase. Pri vytváraní simulácie prostredie umožňuje používateľovi návrh tlačidiel, posuvníkov (umožňujú meniť rýchlosť priebehu simulácie), prepínačov, ukazovateľov, textových polí, poznámok a výstupom. Umožňuje takisto zobrazovať grafy v simuláciách. Modely je možné uložiť ako applety a zabudovať ich do webových stránok.

V Netlogu existujú štyri typy agentov a to sú korytnačky, políčka, spoje a pozorovateľ. Korytnačky sú agenti, ktorí sa pohybujú po svete. Svet je dvojrozmerný a je rozdelený do mriežky s políčkami. Políčka sú oblasť, kde sa pohybujú korytnačky. Spoje sú agenti spájajúci dve korytnačky. Pozorovateľ nie je umiestnený nikde, on sa vlastne pozerá na svet korytnačiek a políčok.

Procedúry sú v Netlogu deklarované prostredníctvom príkazov a reportérov, ktoré jednotlivým agentom hovoria čo majú agenti vykonávať. Príkaz je akcia, ktorú má agent vykonať a reportér vypočíta výsledok a vráti ho. Väčšia časť príkazov sa začína slovesom (napr. create) a reportéri sú podstatné mená alebo frázy. Všetky príkazy a reportéri sú zabudované v tomto nástroji a nazývajú sa primitíva. Každá procedúra má svoj názov, pred ktorým sa nachádza vždy slovo to a celá procedúra je ukončená slovom end. Ak už raz definujeme procedúru v tomto nástroji, tak ju môžeme využiť v celom programe. Niektoré príkazy a reportéri vyžadujú vstupy, čiže hodnoty, ktoré potrebujú na to aby sa vykonala akcia.

Netlogo umožňuje vytvárať rody, čo znamená, že každý definovaný rod sa môže správať v prostredí inak. Konkrétny agent tohto typu je vlastne inštanciu daného rodu – triedy. Rody

korytnačiek definujeme ešte pred samotnými procedúrami na začiatku programu pomocou príkazu breed. Niektoré môžu byť definované aj v množnom čísle. Poradie v akom sú rody deklarované nám určuje v akom poradí budú vrstvy pri zobrazení. Rody definované neskôr sa budú nachádzať na rodoch, ktoré boli definované skôr. Pomocou primitíva set-default-shape určujeme tvary definovaným rodom.

Tlačidlá v Netlogu vytvárame cez prostredie interface a slúžia nám na jednoduché ovládanie modelu. Každý jednoduchý príklad by mal obsahovať tlačidlá pre prípravu prostredia a štart simulácie. Tlačidlo môže slúžiť na spustenie jednorazovej akcie, alebo trvalej akcie zaškrtnutím rámčeka forever. Ak chceme aby tlačidlo spúšťalo iba akcie buď korytnačiek, políčok, alebo spojov, tak to dosiahneme tak, že vytvoríme tlačidlo pozorovateľa a použijeme príkaz ask.

Tvary korytnačiek sú určené vo vektorovom tvare, čiže sú tvorené štvorcami, kruhmi a priamkami a je možné meniť ich veľkosť a takisto ich aj otáčať. Sú uložené v premennej shape a nastavujeme ich pomocou príkazu set. Nastavenie pôvodného tvaru môžeme zmeniť prostredníctvom primitíva set-default-shape. Dajú sa vytvárať aj vlastné tvary prostredníctvom turtle shape editora. Tvary spojov sú určené podobne ako aj u korytnačiek, ale tu editujeme cez link shapes editor a tvary sa skladajú z 0 až 3 čiar ktoré majú rôzny tvar a smer. Spoje majú takisto premennú shape, pôvodný tvar sa mení takisto ako aj u korytnačiek prostredníctvom primitíva set-default-shape.

V mnohých simuláciách čas beží v diskrétnych krokoch (ticks) a počítadlo krokov, ktoré je tu zabudované umožňuje počítať koľko krokov ubehlo v danej simulácii. Aktuálny stav počítadla sa nachádza nad oknom zobrazenia. Reportér ticks načíta aktuálny stav počítadla a príkaz tick zvýši počet o 1. Pomocou príkazu clear-all resetujeme počítadlo na 0, ak potrebujeme vymazať všetko, tak použijeme príkaz reset-ticks. Je dobré používať príkaz tick až, keď všetci agenti vykonajú svoje pohyby a akcie.

Grafy nám v modeloch slúžia na to, aby sme pochopili čo sa v samotných simuláciách deje. Tie sa vytvárajú prostredníctvom panelu interface. Každý graf musí mať svoj jedinečný názov, pretože ho budeme volať v panely procedures. Na zakresľovanie údajov v grafoch sa používajú perá. Pri vytváraní grafu máme na začiatku predvolene vytvorené jedno pero s názvom default. V grafe si môžeme vytvoriť viacero pier. Každé z nich musí mať svoj unikátny názov, pomocou ktorého ho potom budeme volať v panely procedures.

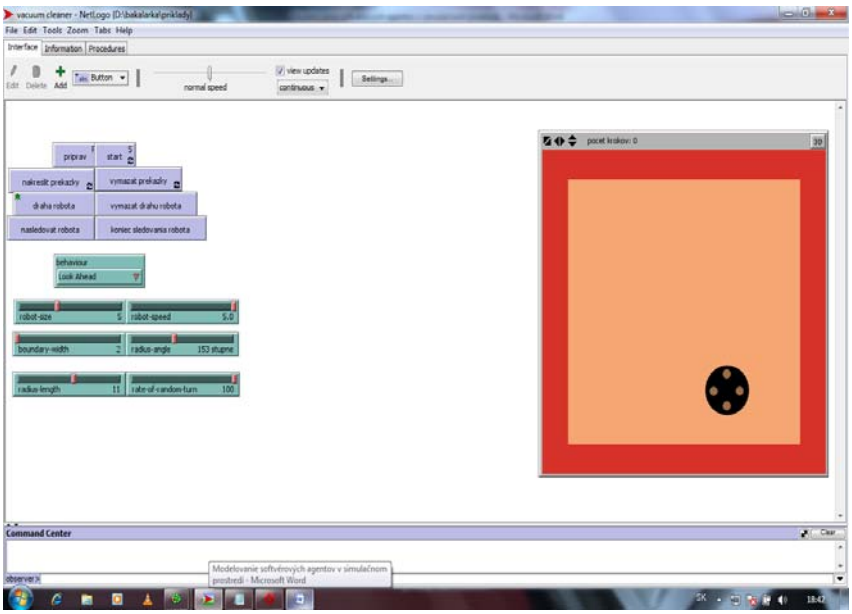
Na vytváranie matíc v Netlogu slúžia maticové dátové štruktúry, ktoré je potrebné zadefinovať na začiatku zdrojového kódu pomocou príkazu extensions [matrix]. Na vytvorenie matice slúži príkaz matrix:make-constant n-rows n-cols number, kde n-rows predstavuje riadky, n-cols stĺpce a number hodnoty vo vytvorenej matici. Pre zápis do matice slúži príkaz matrix:set matrix row-i col-j new-value, kde kde row-i reprezentuje riadky, col-j stĺpce a new-value predstavuje novú hodnotu, ktorá sa zapíše na dané súradnice v matici. Pomocou príkazu matrix:get matrix row-i col-j sa vracajú hodnoty z matice.

Používanie polí je vhodné, ak potrebujeme súbor hodnôt, ktorý má pevne stanovenú veľkosť. Ak poznáme pozíciu položky, tak si ju vieme veľmi rýchlo prečítať, alebo ju zmeniť. Rozšírenie o pole do modelu pridáme tak, že do horného riadnu v paneli procedures pridáme príkaz extensions [array].

IV. NÁVRH JEDNODUCHÉHO AGENTA PRE PREJDENIE ZVOLENÉHO PROSTREDIA

V tejto kapitole popíšeme niekoľko typov jednoduchej implementácie agenta pre prejdenie priestoru (vacuum cleaner agent). V rámci simulácii je možné nakresliť prekážky do tohto simulovaného prostredia, aby bolo možné definovať reálnejšie prostredie. Princíp pohybu agenta je taký, že robot (korytnačka) putuje v prostredí a vyhýba sa prekážkam.

V tomto prostredí sú definované tlačidlá, pomocou ktorých ovládame modelovú simuláciu (viď. Obr.1). Pomocou tlačidla priprav (setup) nastavíme prostredie s vonkajším ohraničením, prostredníctvom spustiť (go) spustíme modelovú simuláciu. Prekážky kreslíme a mažeme prostredníctvom tlačidiel nakresliť prekážky (draw obstacles) a odstrániť prekážky (erase obstacles). Dráhu robota sledujeme a mažeme pomocou tlačidiel dráha robota (toogle plot paths) a vymazanie dráhy robota (erase pats). Poslednými tlačidlami nasledujeme (follow robot) a rušíme nasledovanie robota(stop following robot).



Obr. 1 Rozhranie aplikácie agenta "vacuum cleaner" v prostredí Netlogo.

Pre jednoduché rozlíšenie správania agenta bolo navrhnutých niekoľko typov:

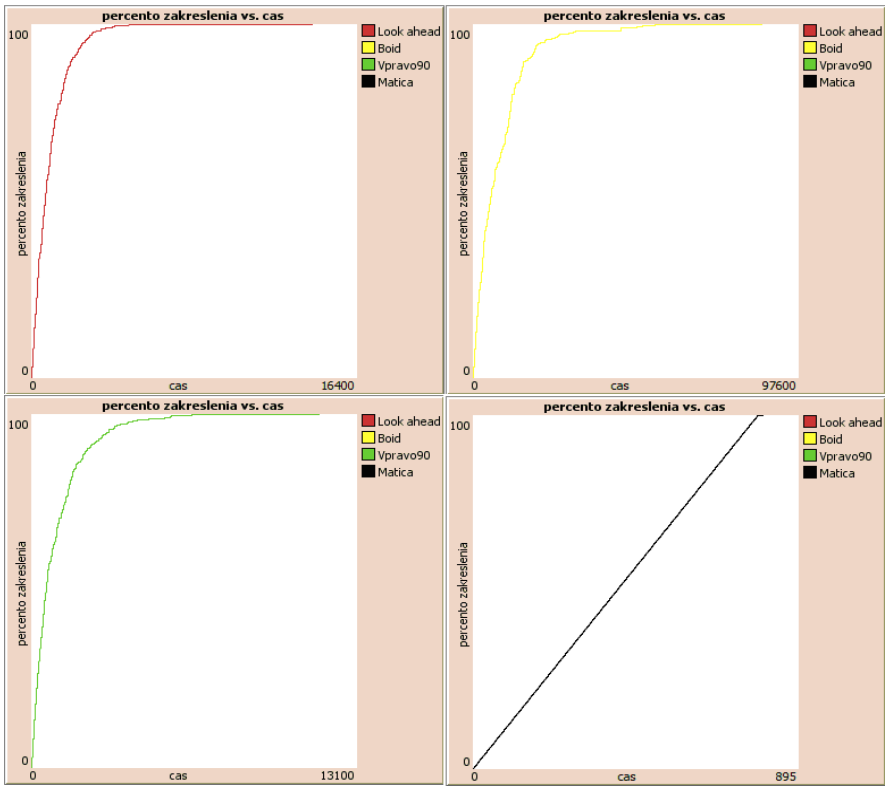
- LookAhead – reaktívny typ, ktorý sa pozerá len pred seba, a ak vidí pred sebou prekážku, tak sa otočí o náhodný uhol a pokračuje ďalej.
- Boid – reaktívny typ, implementujúci tzv. kužeľové videnie, ktorý používa vyhýbanie sa prekážke tak, že sa odráža v pravom uhle doľava.
- Vpravo90 – rovnaký ako typ označený Boid, avšak otáča sa vpravo.
- Matica – typ správania založený na implementácii pamäte zapísanej v podobe matice prostredia zo zápisom navštívených pozícii.

Implementácia prvých troch riešení je v podstate rovnaký, jeho základom je identifikácia prekážky pred agentom a reakcia na to zvoleným spôsobom. Trochu zložitejší model predstavuje agent, ktorý sa učí o prostredí zápisom navštívených miest. Jeho základný princíp pohybu spočíva v tom, že pohyblivý agent (korytnačka) sa pohybuje v prostredí tak, že zisťuje či už niekedy predtým bol na pozíciách okolitých nepohyblivých agentov (políčok). Tento algoritmus spočíva na princípe incidenčnej matice, ktorú reprezentujú jednotliví nepohybliví agenti, čiže v príklade to predstavuje plochu, ktorú môže vysávač povysávať. Ak agent korytnačka, čo v tomto prípade predstavuje vysávač (vacuum clenaer), predtým nikdy nebol na danom políčku, tak hodnota v incidenčnej matici má hodnotu 0. Akonáhle sa vysávač pohne na dané políčko, tak hodnota v incidenčnej matici sa zmení na hodnotu 1. Na začiatku algoritmu je rádius vysávača nastavený na hodnotu 1, čiže overuje 8 políčok okolo seba. Ak medzi nimi nie je taký, ktorého hodnota je 0, tak sa zvýši rádius overovania o 1 a teda už overuje 16 políčok. Ak nájde hodnotu 0, tak sa vysávač pohne, ak nenájde hodnotu 0, tak sa zase zvýši rádius o hodnotu 1 a takto tento algoritmus pokračuje ďalej. Tento postup sa snaží o presnejšie odhadovanie ešte neprehľadného priestoru. Samozrejme, pre obchádzanie prekážky potrebuje takisto procedúru pre reakciu, ktorá môže byť pre jednoduchosť podobná tým, ktoré sú spomenuté pre prvé tri metódy.

Výsledky sú znázornené na Obr.2. Pre jednoduchosť a možnosť porovnania sú tieto vytvárané pre rovnako definované prekážky v priestore. Najmenej úspešný je typ Boid, pretože mal najdlhší čas na zakreslenie prostredia (cca 86500 jednotiek). O niečo úspešnejším bolo správanie typu Look ahead, pretože ten už za výraznejšie kratší časový horizont dokázal zakresliť svoju plochu (cca 14200 jednotiek). Vpravo90, ktorý pri pokusoch dosahoval o niečo lepšie výsledky, je v podstate rádovo na rovnakej úrovni (cca 13 100 jednotiek). Podobnosť týchto dvoch typov je pochopiteľná, vzhľadom k rovnakej stratégii odlíšenej len smerom otáčania. Takisto výrazne lepší čas oproti náhodnému času je pochopiteľný, nakoľko v takom prípade dochádza často k použitiu nevhodného uhlu odrazu reaktívneho agenta.

Oproti jednoduchým reaktívnym agentom bez pamäte bol pre jednoduché porovnanie navrhnutý aj agent s pamäťou. Riešenie s použitím incidenčnej matice navštívenia jednotlivých políčok (v kombinácii s jednoduchou technikou pravouhlého otočenia pre obchádzanie prekážok) zabezpečí bez problémov lineárne vyplnenie priestoru, čo je v podstate optimálne riešenie pre taký jednoduchý problém. Ukazuje tak, že už jeden jednoduchý učiaci prvok môže zmeniť časový graf z asymptoticky lineárneho na lineárny. Dosiahnutý čas pri informovanom

modeli na úrovni cca 800 jednotiek predstavuje zlepšenie o dva rády oproti neinformovaným metódam, predstavuje teda aj úspešnejší typ agenta už pri tak jednoduchom príklade.



Obr. 2 Grafy prejdenia celého prostredia pre rôzne typy správania agentov.

V. ZÁVER

Tento článok popisuje základné pojmy z oblasti agentových systémov, najmä z dôrazom na typy agentov a ich správanie v simulačnom prostredí Netlogo, ako aj praktickú ukážku implementácie agenta pre pohyb v priestore prostredia. V budúcnosti by sme radi využili nadobudnuté znalosti v aplikáciách multiagentových simulácii pre modelovanie viacerých agentov v prostredí v rámci tzv. koalícii spolupracujúcich agentov.

POĎAKOVANIE

Táto práca bola podporená realizáciou projektu Rozvoj Centra informačných a komunikačných technológií pre znalostné systémy (ITMS kód: 26220120030) na základe podpory operačného programu Výskum a vývoj financovaného z Európskeho fondu regionálneho rozvoja.

LITERATÚRA

[1] P. Butka, M. Sarnovský, "Využitie agentových simulácií v socioekonomickej analýze a ich možné rozšírenia," In: Electrical Engineering and Informatics 2, Proceedings of the Faculty of Electrical Engineering and Informatics of the Technical University of Kosice, 2011, pp. 386-391.
[2] B. Frankovič, M. Filka, T.T. Dang, "Multiagentové hierarchické decentralizované systémy," *AT&P Journal*, 2000.
[3] H.S. Nwana, "Software Agents: An overview," *Knowledge Engeneering Review 11(3)*, 1996, pp. 1-40.
[4] H.S. Nwana, L. Lee, N.R. Jennings, "Coordination in Software Agent Systems," *British Telecommunications Technology Journal 14(4)*, 1996, pp. 79–88.
[5] M. Wooldridge, N. R. Jennings, "Intelligent Agents: Theory and Practice," *Knowledge Engineering Review 10(2)*, 1995, pp. 115-152.
[6] P. Vincej, " Modelovanie softvérových agentov v simulačnom prostredí," Bakalárska práca, FEI TU Košice, 2011.
[7] F. Zbořil, " Plánování a komunikace v multiagentních systémech," Dizertační práce, Vysoké učení technické v Brně, 2004.

Implementácia monitorovacieho nástroja pre cloudovú infraštruktúru

Martin SARNOVSKÝ, Peter SASARÁK

Katedra kybernetiky a umelej inteligencie, Fakulta elektrotechniky a informatiky,
Technická univerzita v Košiciach, Slovenská republika

martin.sarnovsky@tuke.sk, peter.sasarak@student.tuke.sk

***Abstrakt* — Táto práca práca sa zaoberá problematikou cloudovej infraštruktúry a vytvorením webovej aplikácie pre monitoring takejto infraštruktúry. V práci popisujeme takúto webovú aplikáciu implementovanú pre prostredie privátneho cloudu pomocou aplikačného frameworku Gridgain.**

***Kľúčové slová* — cloud computing, monitoring**

I. CLOUD COMPUTING

Existuje viacero definícií cloud computingu, no tá najbežnejšia označuje cloud ako sadu spolupracujúcich výpočtových zariadení a služieb, ktoré svoj výkon sprostredkovávajú cez web. Dôležitým aspektom tejto definície sú dynamicky škálovateľné a často virtualizované zdroje poskytované ako služby cez internet na vyžiadanie (tzn. on-demand). Postupne sa cloud computing stáva významným technologickým trendom, a mnoho expertov očakáva, že táto technológia úplne zmení informačno-technologické procesy a IT trh [2]. Cloudová infraštruktúra je konštrukcia, ktorá povoľuje prístup k aplikáciám, ktoré sú v danej chvíli na inom mieste alebo na inom zariadení pripojenom k Internetu. V jednoduchom topologickom zobrazení pozostáva cloud computing z niekoľkých prvkov. Každý prvok má svoj účel a zohráva osobitnú úlohu pri zabezpečovaní základnej cloudovej infraštruktúry: klienti, dátové centrá a distribuované servery. Klienti zohrávajú rovnakú úlohu ako v klasickej lokálnej sieti LAN. Pod klientmi rozumieme koncové zariadenia, ktoré umožňujú užívateľom komunikovať s cloudom a spravovať v ňom svoje informácie. Medzi typických klientov patria: počítače, notebooky, tablety, mobilné telefóny, PDA, atď.

Klientov môžeme rozdeliť do troch kategórií [2] :

- Mobilné zariadenia - do tejto kategórie patria smartphony ako Blackberry, Windows Mobil Smartphone alebo iPhone a taktiež PDA zariadenia.
- Tenkí klienti - tam zahrňujeme počítače, ktoré nemajú interné pevné disky. Server vykonáva všetky operácie, počítače len zobrazia informácie. Stávajú sa čoraz populárnejší keďže majú nižšiu cenu a lepší vplyv na životné prostredie. Majú omnoho nižšie náklady na hardware a energiu. Majú taktiež dlhšiu životnosť. Nie je potrebné ich modernizovanie a nestavajú sa zastarané. Taktiež je tam vysoká bezpečnosť dát vzhľadom k tomu, že dáta sú uložené na serveri a tým pádom je menšia šanca, aby boli stratené, ak sa klientské zariadenie pokazí alebo bude odcudzené. Je tam jednoduchšia oprava alebo výmena. Ak sa klient poškodí, jednoducho sa vymení, a pracovné prostredie sa vráti do stavu, v akom bolo pred zlyhaním.
- Silní klienti - silný klient je bežný počítač, ktorý je pripojený ku cloudu pomocou webového prehliadača.

Dátové centrum je zberňa serverov, kde sú umiestnené aplikácie, na ktoré sa užívateľ prihlasuje. Rastúcim trendom v IT svete je virtualizácia serverov. To znamená, že softvér dovoľuje viacero inštancií virtuálnych serverov, ktoré sa budú používať. Týmto spôsobom môžeme mať desiatky virtuálnych serverov bežiacich na jednom fyzickom serveri. Počet virtuálnych serverov, ktoré môžu existovať na jednom fyzickom serveri, závisí na veľkosti a

rýchlosti fyzického servera a na tom, aké aplikácie na ňom budú bežať.
Nie je možné aby boli všetky servery umiestnené na jednej lokalite. Veľmi často sa stáva, že sú umiestnené v geograficky odlišných miestach. Vďaka cloudu sa však správajú ako keby boli bezprostredne vedľa seba. To dáva poskytovateľovi služieb väčšiu flexibilitu a bezpečnosť. Napríklad Amazon má svoje servery po celom svete. Ak by sa vyskytlo zlyhanie na jednom mieste, k službám by sme mohli stále pristúpiť pomocou iného webu. Takisto ak cloud potrebuje väčšiu hardvérovú kapacitu, môže ju pridať z cloudu na inom mieste.

II. MODELY POSKYTOVANIA SLUŽIEB

Softvér ako služba (SaaS, Software as a Service) je model, v ktorom žiadosť hosťuje ako služba zákazníkom, ktorí majú prístup na internet. Hlavnou ideou je, že zákazník môže použiť softvér bez toho, aby si musel zakúpiť licenciu. Poskytovateľ takýchto služieb má na starosti všetky opravy, modernizácie, ako aj udržanie plynulosti chodu infraštruktúry. Existuje veľa druhov softvérov, typický je ten, ktorý vykonáva jednoduchú úlohu bez veľkej potreby komunikovať s ostatnými systémami. Model SaaS je najvhodnejší pre zákazníkov, ktorí nie sú ochotní investovať do neustáleho vývoja softvéru a jeho modernizácie, no na druhej strane potrebujú využívať vysoko výkonné aplikácie. Softvér je riadený na centrálnom mieste, teda zákazníci majú neustály prístup ku svojim aplikáciám. Jedinou podmienkou je prístup na internet [3]. PaaS (Platforma s a Service) dodáva všetky zdroje potrebné na vytváranie aplikácií bez toho, aby sa museli stiahnuť a nainštalovať. PaaS služby v sebe zahŕňajú návrh aplikácie, vývoj, testovanie, nasadenie a hosting. PaaS tiež podporuje vývoj webových aplikácií ako sú Simple Object Access Protocol (SOAP) a Representational State Transfer (REST), ktoré umožňujú výstavbu viacerých webových služieb. Užívateľ nemôže kontrolovať infraštruktúru, má len kontrolu nad nasadenou aplikáciou. PaaS ponúka podporu a pomoc na vytváranie užívateľských rozhraní. Rozhranie je zvyčajne založené na HTML a JavaScripte. Keďže sa očakáva využívanie služieb viacerými používateľmi súčasne, PaaS je navrhnutý tak, aby poskytoval automatické zariadenia pre súbežnosť správy, škálovateľnosť a zabezpečenie [3]. Posledným modelom je IaaS (Infrastructure as a Service) a umožňuje vytvoriť infraštruktúru v cloude, pričom ponúka vysokú škálovateľnosť. Odvoláva sa na výpočtové zdroje ako službu. To zahŕňa virtualizované počítače s garantovaným výpočtovým výkonom a vyhradenou šírkou pásma pre ukladanie a prístup na internet [3] .

III. IMPLEMENTÁCIA NÁSTROJA PRE MONITORING ÚLOH

Našou úlohou bolo vytvoriť rozhranie pre monitoring jednotlivých uzlov pripojených do distribuovaného prostredia privátneho cloudu. Cieľom bolo získať údaje pripojeného užívateľa a získať a vizualizovať prostredníctvom webovej aplikácie dôležité a zaujímavé informácie o hardvérových parametroch jednotlivých výpočtových uzlov pripojených do cloudu. Rozhranie bolo vyvíjané vo vývojovom prostredí Eclipse, ktoré je určený na programovanie v jazyku Java a C++. V našom prípade sme zvolili jazyk Java a voľne dostupný framework Gridgain, ktorý poskytuje API pre programovanie Java cloudových aplikácií.

Gridgain je cloudový aplikačný framework, ktorý umožňuje vývoj vysoko škálovateľných distribuovaných aplikácií [4]. Podporuje implementáciu v programovacích jazykoch Java a Scala. Gridgain podporuje implementáciu pre výpočtové gridy a dátové gridy, má implementované metódy distribuovaného počítania (MapReduce). Našim cieľom bolo vytvoriť monitorovaciu aplikáciu pre privátnu cloudovú infraštruktúru, na ktorá je postavená a využívaná pre distribuované Java Gridgain aplikácie.

Cieľom bolo navrhnúť webovú aplikáciu, ktorá registruje každý uzol, ktorý ponúka svoj diskový priestor a výpočtový výkon. Aplikácia by mala byť schopná vizualizovať všetky pripojené uzly a k nim dostupné relevantné informácie. Ku každému uzlu vieme prostredníctvom metód Gridgain API pristúpiť k týmto dátam:

- Id uzla
- Názov počítača
- Názov užívateľa
- Java verzia
- Gridgain verzia
- Operačný systém
- Počet procesorov

- Bežiace úlohy
- Voľné miesto na disku
- Veľkosť pamäte
- Použitá pamäť
- IP adresa
- Id úlohy (ak sa vykonáva)

Na získanie dát boli implementované metódy využívajúce Gridgain API. Aplikácia k nim pristupuje metódou POST umiestnenou v HTML formulári. Výsledky sú umiestnené v tabuľke. Na základe hardvérových komponentov si môže užívateľ vybrať, ktoré uzly budú spracovávať jednotlivé úlohy. Využitím metódy POST aplikácia vizualizuje rôzne hardvérové údaje ako napríklad: Operačný systém, počet procesorov, voľné miesto na disku, veľkosť pamäte, použitá pamäť atď. Pri každom načítaní stránky servlet pristupuje ku API Gridgain a získava údaje a vykonáva úlohy. Keďže bežný webový server nedokáže spracovať servlety, museli sme použiť servlet kontajner Tomcat. Na spracovanie údajov zo stránky využívame metódu POST, ktorá je umiestnená v HTML formulári. V našom prípade využívame na odosielanie údajov metódu Post, no v servlete iba metódu doGet, preto parametre Post smerujeme do doGet. Gridgain nám umožňuje získať informácie o všetkých nodoch. Výpis tabuľky je cyklus, ktorý prechádza všetkými nodmi a generuje jednotlivé riadky v tabuľke, následne sú vypísané v rozhraní. Pri výpise jednotlivých nodov získavame informácie, napr. o veľkosti voľného miesta na disku. Tieto hodnoty sú však uvádzané v bytoch. Pre jednoduchosť zobrazenia pre užívateľa využívame funkciu, ktorá prevádza túto hodnotu na vyššie jednotky.

```
for (GridNode node : G.grid().nodes()) {

htmlout.println("<tr><td colspan='13'>Nod ID:<strong>" + node.id().toString()+"</strong> spusteny
"+format.format(node.metrics().getStartTime())+"</td></tr>");

        htmlout.println("<tr>");

        htmlout.println("<td>" +node.attribute("USERDOMAIN")+"</td>");

        htmlout.println("<td>" +node.attribute("user.name")+"</td>");

        htmlout.println("<td>" +node.attribute("java.specification.version")+"</td>");

        htmlout.println("<td>" +node.attribute("org.gridgain.build.ver")+"</td>");

        htmlout.println("<td>" +node.attribute("os.name").toString()+"</td>");

        htmlout.println("<td>" +node.attribute("os.arch")+"</td>");

        htmlout.println("<td>" +node.metrics().getAvailableProcessors()+"</td>");

        htmlout.println("<td>" + node.metrics().getCurrentActiveJobs()+"</td>");

        htmlout.println("<td>" + PrevodBytov(node.metrics().getFileSystemFreeSpace())+"</td>");

        htmlout.println("<td>" +
      PrevodBytov(node.metrics().getHeapMemoryMaximum())+"</td>");

        htmlout.println("<td>" + PrevodBytov(node.metrics().getHeapMemoryUsed())+"</td>");

        htmlout.println("<td>" + node.externalAddresses().toString()+"</td>");

        htmlout.println("</tr>");
```

Obr. 1 Typy atribútov vizualizovaných v rámci monitorovacej aplikácie.

Výsledná webová aplikácia a jej používateľské rozhranie je zobrazené na Obr. 2. Obsahuje informácie o všetkých uzloch zapojených do distribuovanej infraštruktúry a rovnako aj implementuje metódu pre umožnenie vzdialeného odpojenia uzlov cez monitorovacie rozhranie.

GridGain HTML

[Aktualizovat obsah]

Nazov pocitaca	Nazov uzivatela	Java ver.	GG ver.	OS	Arch.	Pocet proc.	Beziace ulohy	Volne miesto na disku	Velkost pamate(VM)	Pouzita pamat	IP adresa	Uloha
Nod ID:**d3ef8594-c3f9-4fbb-b332-c7e902ceb8b4** spusteny 17.05.2011 10:21.42												
sasi-PC	sasi	1.6	3.0.5c-25012011	Windows 7	x86	2	0	61 GB	259 MB	7 MB	[147.232.180.86]	N/A (lokalny nod)
Nod ID:**9cab57e6-8c65-463c-9396-827a0f90eecf** spusteny 17.05.2011 10:22.53												
MAROS234	Maros	1.6	3.0.5c-25012011	Windows 7	x86	2	0	30 GB	268 MB	21 MB	[147.232.180.88]	G.stop()
Nod ID:**5d60c799-0d0d-4789-a626-129421f11149** spusteny 17.05.2011 10:23.21												
Sisa-PC	Sisa	1.6	3.0.5c-25012011	Windows Vista	x86	2	0	77 GB	268 MB	20 MB	[147.232.181.36]	G.stop()
Nod ID:**dbbc2ca4-7029-43b2-966a-83908f008195** spusteny 17.05.2011 10:23.20												
Arthys-PC	Arthys	1.6	3.0.5c-25012011	Windows 7	x86	4	0	384 GB	268 MB	21 MB	[147.232.180.89]	G.stop()

Obr. 2 Typy atribútov vizualizovaných v rámci monitorovacej aplikácie.

IV. ZÁVER

Článok popisuje vytvorenie rozhrania pre monitoring úloh, ktoré sú vykonávané na cloudovej infraštruktúre. Využili sme vývojové prostredie Eclipse, framework Gridgain a Servlet kontajner Tomcat. Hlavným cieľom bolo implementovať základnú funkcionalitu monitorovacieho rozhrania pre privátny cloud, ktorá bude v budúcnosti rozširovaná. Cieľom je implementovať komplexné administrátorské rozhranie pre správu cloudového prostredia pre riešenie distribuovaných úloh a aplikácií postavených na platforme Gridgain.

POĎAKOVANIE

Táto práca bola podporená realizáciou projektu Rozvoj Centra informačných a komunikačných technológií pre znalostné systémy (ITMS kód: 26220120030) na základe podpory operačného programu Výskum a vývoj financovaného z Európskeho fondu regionálneho rozvoja.

LITERATÚRA

[1] HALL, Marty. Java servlety a stránky JSP. [s.l.] : Neocortex spol. s.r.o.,2011, 2000. 586 s. ISBN 80-86330-06-0B. Frankovič, M. Filka, T.T. Dang, "Multiagentové hierarchické decentralizované systémy," *AT&P Journal*, 2000.

[2] VELTE, Anthony T.; VELTE, Toby J.; ELSENPETER, Robert. Cloud computing : A practical Approach. The United States : The McGraw-Hill Companies, 2010. 334 s. ISBN 978-0-07-162694-1.Handbook of cloud computingH.S. Nwana, L. Lee, N.R. Jennings, "Coordination in Software Agent Systems," *British Telecommunications Technology Journal 14(4)*, 1996, pp. 79─88.

[3] FURHT, Borko; ESCALANTE, Armando. Handbook of cloud computing. 233 Spring Street, New York, USA : Springer, 2010. 653 s. ISBN 978-1-4419-6523-3.P. Vincej, " Modelovanie softvérových agentov v simulačnom prostredí," Bakalárska práca, FEI TU Košice, 2011.

[4] Http://www.gridgainsystems.com [online]. 14.7.2007 [cit. 2011-05-18]. 10 reasons to use gridgain. Dostupné na internete : <http://www.gridgainsystems.com/wiki/display/GG15UG/10+Reasons+to+Use+GridGain>.

Luwak: System na extrakciu informácii z homogénnych textových dokumentov

Peter SMATANA

Katedra kybernetiky a umelej inteligencie, FEI TU v Košiciach, Slovenská republika

psmatana@gmail.com

Abstrakt – **Článok pojednáva o aplikovanom výskume v rámci projektu ITLIS, ktorého cieľom je vyvinúť vyhľadávací a vizualizačný nástroj nad dátami integrovanými z rozličných zdrojov. Zdrojom dát sú verejne prístupné registre obsahujúce informácie hlavne o ľudoch, firmách a iných organizáciach. V článku je prezentovaný nekonvenčný prístup k extrahovaniu dát z textových dokumentov. Tento prístup je založený, na predpoklade, že dokumenty nie sú štrukturované no popisujú rôzne subjekty zo špecifického uhla pohľadu. Tento popis býva na vysokej odbornej úrovni (napr. záznamy o pacientoch, rozhodnutia sudcov vo veciach obchodných spoločnosti), ktorý je laikom ťažko zrozumiteľný no vyznačuje sa množstvom stabilných formulácii. Tieto formulácie systém Luwak identifikuje a následne transformuje do štrukturovanej formy pre potreby presného vyhľadávania a uvažovania nad spracovanými dokumentami. Vrámci vývoja systému bol navrhnutý algoritmus na určovanie vzdialenosti dvoch reťazcov viacvrstvová Levenshteinova vzdialenosť, ktorý je základom pre objavovanie nových vzorov.**

Keywords – **extrakcia informácií, rozpoznávanie názvoslovných entít, viacvrstvová Levenshteinova vzdialenosť, spracovanie prirodzeného jazyka**

I. Úvod

Google[1] je synonymom vševeda. V dnešnej dobe je odpoveďou na každú otázku, aj keď v skutočnosti je to len systém, ktorý veľmi precízne indexuje veľké množstvo dát, ktoré sú na webe a tým uľahčuje prácu mnohým ľudom pri riešení ich každodenných problémov. Rýchly prístup k informáciám je dôležitý hlavne v krízových situáciach či už pri uzatváraní obchodu a preverenia si svojich obchodných partnerov alebo pri určovaní diagnózy pacientov na základe podobných príznakov u predchádzajúcich pacientoch ako aj v mnohých iných prípadoch. Tieto dáta sú zvyčajne v neštrukturovanej forme a pri štandardnom full-textovom vyhľadávaní (na ktorom je postavený aj Google vyhľadávač) je nemožné získať relevantné informácie. Preto vznikajú extrakčné systémy, ktoré transformujú dokumenty do tabuľkovej formy, v ktorej je následne možné presné vyhľadávanie pomocou rôznych dopytovacích jazykov ako aj inferencia nových informácii.

Popisovaný systém je časťou projektu ITLIS[2], ktorého cieľom je vyvinúť vyhľadávací a vizualizačný nástroj nad dátami integrovanými z rozličných zdrojov. Zdrojom dát sú verejne prístupné registre obsahujúce informácie hlavne o ľudoch, firmách a iných organizáciach. Mnohé registre obsahujú homogénne[3] neštrukturované dáta v textovej podobe, ktoré je potrebné transformovať do tabuľkovej formy. Túto úlohu na seba preberá semi-automaticky anotačný systém Luwak[4] vyvíjaný v rámci tohto projektu. V nasledujúcich odsekoch bude predstavená architektúra systému ako aj algoritmus na určenie vzdialenosti dvoch reťazcov viacvrstvová Levenshteinova vzdialenosť (Multilayer Levenshtein distance - MLD).

II. Extrakcia informácii

Ak sme postavený pred úlohu spracovania textových dokumentov, tak sa sa nachádzame na križovatke mnohých ciest, nazývanej spracovanie prirodzeného jazyka. Pod týmto termom rozumieme skupinu úloh, ktoré riešia jej dielčie úlohy (získavanie informácii, extrakcia informácii,

[1] http://google.com

[2] http://itlis.eu

[3] pod homogénnou textovou množinou budeme rozumieť dokumenty, ktoré popisujú rôzne subjekty z rovnakého uhla pohľadu, čiže obsahom a formuláciami viet sú si veľmi podobné no líšia sa v identifikátoroch, hodnotách, ktoré daný subjekt pre určitú vlastnosť nadobúda

[4] slovo luwak označuje v indonézčine zviera - cibetku, ktorá se živí hmyzom a tropickými plodmi, taktiež aj plodmi kávovníka, z ktorých sa vyrába najdrahšia káva na svete. Tu je možné vidieť paralelu, kde popisovaný anotačný systém sa snaží označiť časti textu s vysokou informačnou hodnotou.

zodpovedanie otázok, klasifikácia textových dokumentov, zhlukovanie textových dokumentov a.i.). Spracovanie prirodzeného jazyka momentálne nie je v ľudských silách riešiť komplexne, preto táto križovatka ciest a je na znalostnom inžinierovi aby si zvolil tú správnu.

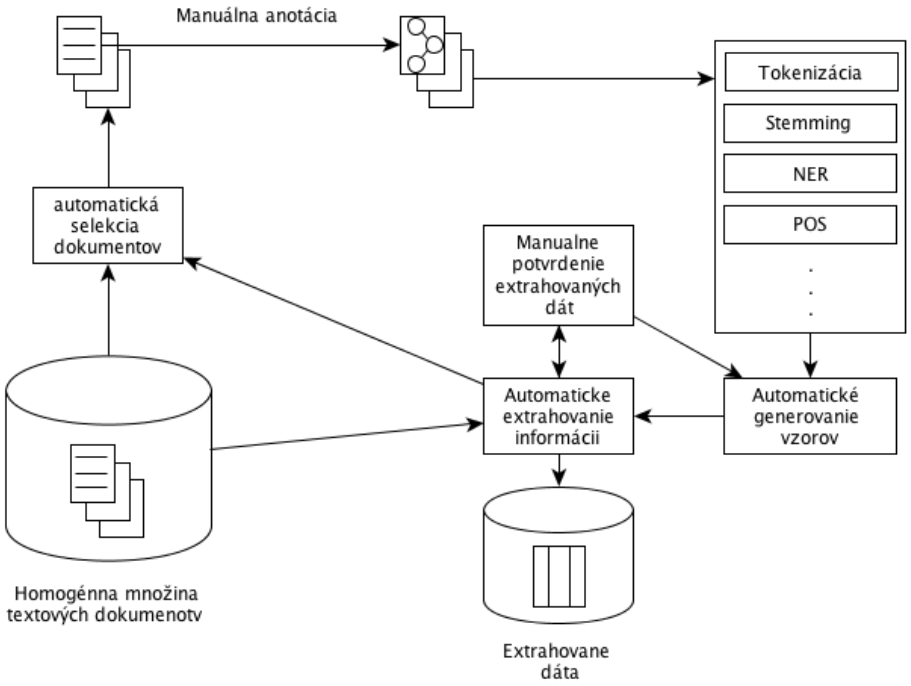
Úloha, ktorú sme sa podujali riešiť zapadá do rámca extrahovania informácii z textových dokumentov. Jedna z mnohých definícii, ktorá ho vystihuje je: *Extrakcia informácii je identifikácia a následná klasifikácia a štruktúrovanie špecifických informácií nájdených v neštrukturovaných dátových zdrojoch (ako sú napr. texty v prirodzenom jazyku) do sémantických kategórií tak, aby boli pripravene pre ďalšie metódy spracovania informácii.[1]*

Bolo vynájdených mnoho metód pre IE, v počiatkoch to boli hlavne manuálne tvorené pravidlové systémy[2] neskôr začali dominovať systémy postavené na algoritmoch strojového učenia[3][4][5], tie však stále vyžadujú kvalitne anotované veľké trénovacie množiny. Z týchto dôvodov sa tieto metódy kombinujú s prvkami aktívneho a semi-kontrolovaného učenia, ktoré pomáhajú zmenšiť veľkosť trénovacej množiny. Tieto prvky používame aj v našom systéme, ktorého návrh je popísaný v nasledujúcej kapitole.

III. Návrh systému

Systém Luwak je navrhnutý ako interaktívny systém, ktorý je v spolupráci s doménovým a znalostným expertom schopný transformovať neštrukturované textové dokumenty do štrukturovanej formy. Extrakčný proces môžme rozdeliť do dvoch fáz:

- Prípravná
  - Oboznámenie sa s homogénnou množinou dokumentov
  - Definovanie cieľa doménovým expertom
  - Príprava množiny dokumentov
  - Prvotný návrh annotačného modelu
  - Výber a konfigurácia modulov predspracovania textov
- Samotná extrakcia - iteratívny proces (viď Obr. 1)



Obr. 1 Schéma extrakčného procesu systému Luwak

V prvej fáze je nevyhnutná komunikácia doménového experta (ako používateľa extrahovaných dát) a znalostného inžiniera, ktorý vyhodnotí technické možnosti a riadi celý proces spracovania dokumentov. Samotná extrakcia je viac menej manuálna práca, v ktorej systém riadi tvorenie trénovacej množiny a presnosť a návratnosť objavených extrakčných vzorov. Iteračný proces končí dosiahnutím úplného pokrytia vstupnej množiny dokumentov alebo jej definovanej časti.

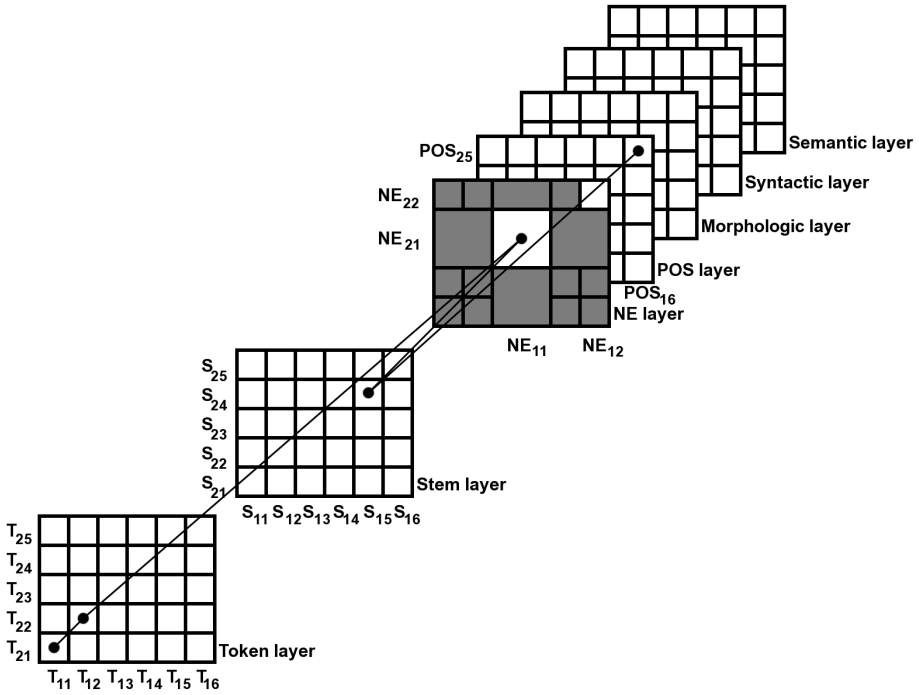
Nie je možné sa vyhnúť manuálnej tvorbe trénovacej množiny ale pomocou znalosti z teórie učenia môžeme minimalizovať množstvo ručne anotovaných dokumentov. To je aj cieľom bloku automatickej selekcie dokumentov v kooperácii s blokom automatického extrahovania informácii.

IV. Viacvrstvová Levenshteinova vzdialenosť

V teórií informácii je Levenshteinova vzdialenosť[6] metrika nad reťazcami na meranie počtu rozdielov medzi dvomi sekvenciami. Levenshteinova vzdialenosť medzi dvomi reťazcami je definovaná ako minimálny počet zmien potrebných na transformovanie jedného reťazca na druhý s povolenými operáciami vloženie, zmazanie a substitúcia jedného znaku. My sme ju modifikovali na určenie podobnosti dvoch sekvencii tokenov s tým, že pri porovnávaní sekvencii sa uvažuje aj o možných anotáciach na vyšších úrovniach ako sú: rozpoznaných pomenovaných entít, syntaktická, sémantická, atď. Na Obr. 2 je vidieť ako sa môžu napríklad porovnať nasledujúce dve vety.

- Spoločnosť prenajíma Slovenskej sporiteľni predmetné BMW.
- Spoločnosť prenajíma Prvej správcovskej auto.

Ak zoberime do úvahy, že vo fáze predspracovania boli rozpoznané pomenované entity: Slovenská sporiteľňa, BMW a Prvej správcovskej; POS tagger nám okrem iných anotácii BMW a auto označil ako podstatne meno (v tomto prípade by bola vhodnejšia a potrebná syntaktická úroveň čiže anotáciu predmet), tak vzdialenosť pomocou Levenshteinovej vzdialenosti by bola 4 no Viacvrstvová Levenshteinova vzdialenosť pri porovnaní reťazcov vrátila hodnotu 1 čo vystihuje podobnosť daných viet.



Obr. 2 Ukážka porovnania dvoch kontextov pomocou MLD

Nasledujúci pseudokód ukazuje ako je možne implementovať viacvrstvovú Levnshteinovu vzdialenosť. Porovnávame slovo s dĺžky n a slovo t dĺžky m, kde obe slová obsahujú p anotačných vrstiev.

```
int MLDistance(char s[1..n, 1..p], char t[1..m, 1..p]) {
    declare int MLD[0..n, 0..m, 1..p]
    declare int i, j, k cost

    for i := 1 to m
        MLD[0, i] := infinity
    for i := 1 to n
        MLD[i, 0] := infinity
    MLD[0, 0] := 0

    for i := 1 to n
        for j := 1 to m
            for k := 1 to p
            cost:= d(s[i, k], t[j, k])
            MLD[i, j, k] := cost + minimum(MLD[i-1, j   , 1..p],
```

```
                                        MLD[i   , j-1, 1..p],
                                        MLD[i-1, j-1, 1..p])

    return minimum(MLD[n, m, 1..p])
}
```

V. ZÁVER

I keď analýza neštrukturovaných dokumentov v prirodzenom jazyku je veľmi závislá na kvalite predspracovania pomocou lingvistických metód a veľkosti trénovacej množiny, tak na základe obmedzenia množín dokumentov na homogénne sady očakávame aj so základnými modulmi na predspracovanie textových dokumentov a použitím techník aktivného a semi-kontrolovaného učenia dosiahnutie vysokej kvality extrahovaných dát.

Rýchle a efektívne porovnávanie jednotlivých kontextov anotácii je možné aj vďaka navrhnutému MLD algoritmu, ktorý umožňuje porovnávať reťazce cez viacero anotačných vrstiev, čo je nevyhnutné pre správnu funkčnosť extrakčných algoritmov.

VI. POĎAKOVANIE

Rad by som vyjadril svoje poďakovanie prof. Ing. Jánovi Paraličovi, PhD a svojím kolegom Ing. Petrovi Bednárovi, PhD. a Ing. Petrovi Kostelníkovi, PhD. za cenné rady a pripomienky pri vývoji systému Luwak.

LITERATÚRA

[1] M.-F. Moens, *Information Extraction: Algorithms and Prospects in a Retrieval Context (The Information Retrieval Series)*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2006.

[2] G. Dejong, "An overview of the FRUMP system," in *Strategies for natural language processing*, W. G. Lehnert and M. H. Ringle, Eds. Hillsdale, NJ: Lawrence Erlbaum, 1982.

[3] *University of Massachusetts : Description of the CIRCUS System as Used for MUC-4*, ser. FOURTH MESSAGE UNDERSTANDING CONFERENCE (MUC-4), Proceedings of a Conference Held in McLean, Virginia, June 16-18, 1992, 1992.

[4] T. R. Leek and T. R. Leek, "Information extraction using hidden markov models," 1997.

[5] S. Miller, M. Crystal, H. Fox, L. Ramshaw, R. Schwartz, R. Stone, R. Weischedel, and T. A. Group, "Algorithms that learn to extract information bbn: Description of the sift system as used for muc-7," in *In Proceedings of MUC-7*, 1998.

[6] V. Levenshtein, "Binary Codes Capable of Correcting Deletions, Insertions and Reversals," *Soviet Physics Doklady*, vol. 10, p. 707, 1966.

Comparison of MF ARTMAP Neural Network and MF ARTMAP Ceural Network with Modification of Fuzzy Relation

Peter SMOLÁR, Peter SINČÁK

Department of Cybernetics and Artificial Intelligence, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Slovak Republic

peter.smolar@tuke.sk, peter.sincak@tuke.sk

***Abstract* — This paper deals with MF ARTMAP neural network, which belongs to family of ARTMAP neural networks. This network combines in own structure Adaptive Resonance Theory (ART) in learning process and Fuzzy Logic in the output. We propose modification of fuzzy relation for achieving of better results in learning of very complex data in the classification domain. We were inspired in proposal of new equation of fuzzy relation by equation of multivariate normal probability distribution. At the close of the article we compare our modification with classical MF ARTMAP neural network.**

***Keywords* — MF ARTMAP neural network, ART theory, fuzzy logic**

I. INTRODUCTION

Hand in hand with the expansion of the application potential of the information technology, raises demands for fast processing of large amounts of data. This is the reason why methods used for obtaining relevant information asked by customers are under pressure.

Artificial intelligent (AI) methods are one of the possible ways how to handle with data. Neural networks, among others AI techniques, posses effectives features: the ability to adapt to dynamically changing problems, the ability to learn and also to handle uncertainty.

This paper introduces one of the type of NN, the network class ART (Adaptive Resonance Theory) and specifically ARTMAP. The approaches based on ART or ARTMAP technology are rather flexible and much more suitable for these situations when class consists of various clusters in the feature space.

This paper is organized as follows: Section II theoretically talks about the ART-like neural networks. The basic description of MF-ARTMAP is given. Section III contains our proposal of the improvement of this type of NN. Section IV presents results of comparison between classical MF ARTMAP neural network and MF ARTMAP network with modification of fuzzy relation.

II. NEURAL NETWORKS OF TYPE ARTMAP

Adaptive Resonance Theory (ART) was developed by Grossberg and Carpenter [1]. It was biologically inspired, trying to imitate how the brain processes information –analysis of human cognitive information and stable coding in a complex input environment [2]". Simply described, the pattern matching process compares the current input with a learnt category representation. Then this process leads to a resonant state (focusing attention and triggering category learning) or to a self-regulating parallel memory search leading to a resonant state. The important feature of ART NN is that if an established category is selected, the category´s representation may be refined to incorporate new information from the current input. In case that the search ends by selecting an antecedently untrained node, the ART NN establishes a new category.

ART NN has been used in a wide spectrum of application fields, including problems of learning and classification, recognition and prediction. Different types of ART NN have been

developed: the first one by Grossberg was an unsupervised learning system used for categorize binary input patterns; the next ones were extended to categorize also the analog input patterns.

Neural networks of type ARTMAP, known also as predictive ART, have supervised learning structure, with ability to self-organize arbitrary mappings from input vectors representing features such as spectral values and terrain variables, to output vectors, representing predictions such as vegetation classes in a remote sensing application. If fuzzy ART units are used, we talk about Fuzzy ARTMAPS. It calculates the membership function of the point from the feature space to the fuzzy class. MF ARTMAP is based on the assumption that data in feature space are organized in fuzzy clusters. Grossberg in [1] it describes as a synthesis of fuzzy logic and adaptive resonance theory. Figure 1 illustrates the basic topology of the MF ARTMAP NN and in figure 2 is illustrated membership function of fuzzy relation for one cluster.

I. ANALYSIS OF MF ARTMAP NEURAL NETWORK

We analyzed MF ARTMAP neural network during the phase of learning and classification, we found some "weak points" of the network. One of these deficiencies can be described as small possibility to adapt on very complex input data. This characteristic is caused by used equation of fuzzy relation and his form is:

$$A(\bar{x}_i) = \int_U \frac{1}{1 + \left| \frac{(\bar{x}_{gk} - \bar{x}_i)^2}{E_k} \right|^{F_k}} / \bar{x}_i \qquad (1)$$

where index m is number of dimensions input space, $j=1, 2, \ldots, m$ is index of dimension of input space, $U = X_1 \times X_2 \times \ldots \times X_m$, represents quotation of input space, index n is count of all samples in training set, $i = 1,2,\ldots,n$ is index of current sample, index p is number of momentary created fuzzy clusters, $k = 1,2,\ldots,p$ is index of current fuzzy cluster, denotation $A(\bar{x}_i)$ is membership of sample $\bar{x}_i$ to fuzzy cluster k, vector $\bar{x}_i$ is arbitrary point in input space, vector $\bar{x}_{gk}$ represents center of fuzy cluser k, E_k and F_k are parameters of fuzzy cluster k.

This equation dictates change of shape of cluster. It is showed in the figure 1.

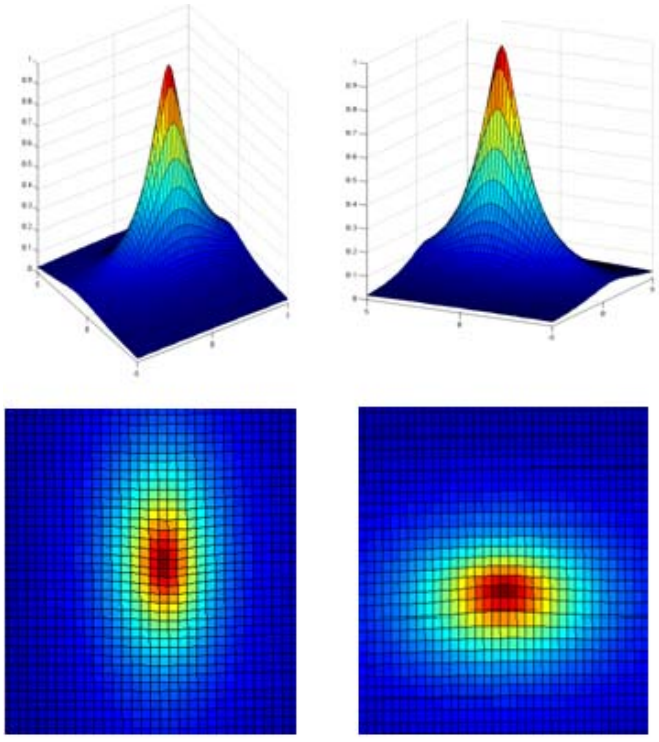


Fig. 1 Shapes of equation of fuzzy relation [6].

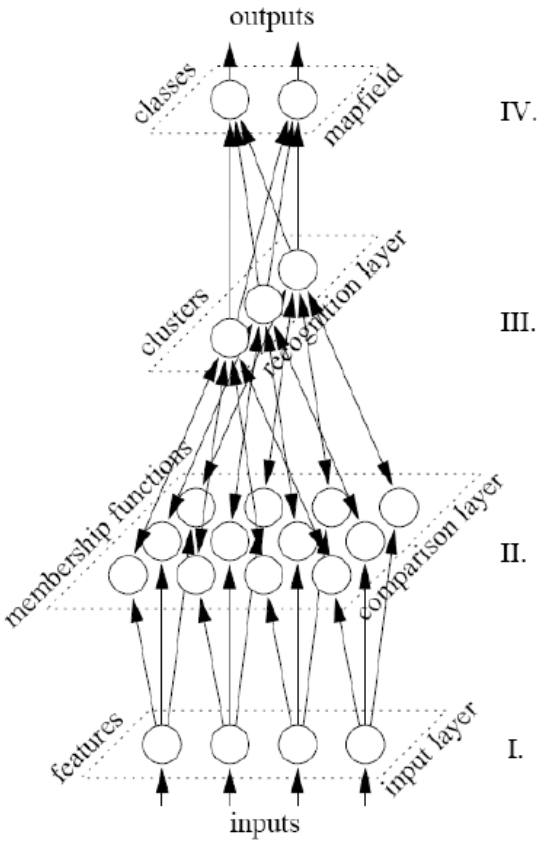


Fig. 2. Basic topology of the MF ARTMAP neural network [6].

Based on this analysis we propose modification of MF ARTMAP neural network, where we expect increase of the accuracy of classification results.

II. Proposal of Equation of new Fuzzy Relation

We want replace equation of classical fuzzy relation with equation of multivariate normal (Gauss) probability distribution, where covariance matrix knows better to adapt on complex data.

To ensure range of membership value of fuzzy relation between <0,1>, part of equation before exponent is skipped. Form of equation of fuzzy relation is defined:

$$A(\vec{x}) = e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)} \quad (2)$$

where Σ is covariance matrix of input samples in feature space and x is vector of values of input sample and is vector of average values.

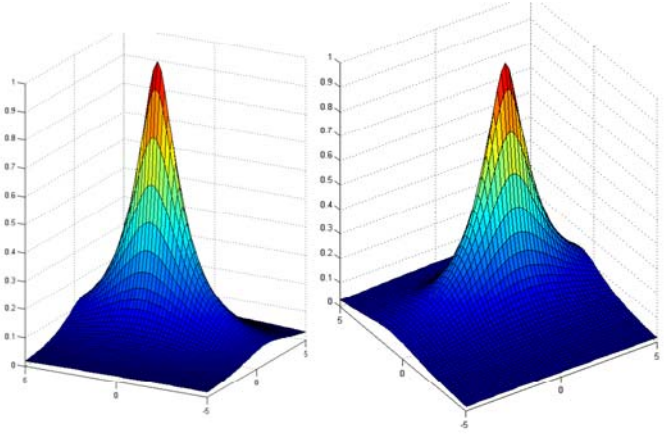


Fig. 3 Shape of modified fuzzy relation.

III. EXPERIMENTS WITH TESTED IMAGES

We performed experiments with MF ARTMAP neural network and MF ARTMAP neural network with modified equation of fuzzy relation.
These comparisons we did with two images. First image is "circle in the square" (Fig. 4, left side) and second one is the "spiral" (Fig. 4, right side).



Fig. 4 Tested images. On the left is tested image "Circle in the square" and on the right is image "Spiral".

The experiments we divided into several parts. First, we tested classical MF ARTMAP network, and we investigated the most optimal parameters of the network. Then we investigated the modified MF ARTMAP neural network the same way. Finally, we evaluated and compared these results between two concepts of various equations of fuzzy relation.

A. Classical MF ARTMAP neural network

We permanently adjusted parameters of neural network to following values:

- Parameter F = 0.6,
- Parameter E = 0.5,
- Minimal membership threshold = 0.5.

In the case, tested image was "Circle in the square", we changed count of samples in range from 10% to 50% of the all samples, used in classification. Sample step was set to 10. Maximal distance threshold we changed in range from 0.05 to 0.35 in step 0.05.

We found out, that the best results were achieved with 1,85% classificatory error, with using 50% of the all samples in learning and recognition layer threshold was set to 0.05. Number of created samples was 218.

In the case, tested image was "Spiral", we changed count of samples in range from 10% to 50% of the all samples. Sample step was set to 10. Maximal distance threshold we changed in range from 0.03 to 0.21 in step 0.03.

We found out, the best results were achieved with14.67% classificatory error, using 30% of the all samples in learning and recognition layer threshold was set to 0.03%. Number of created samples was 652.

B. Modified MF ARTMAP neural network

We permanently adjusted minimal membership threshold of neural network to values 0.5.

In the case, tested image was "Circle in the square", we changed count of samples in range from 10% to 50% of the all samples, used in classification. Sample step was set to 10. Covariance range we changed in range from 0.5 to 2.3 in step 0.3.

We found out, that the best results were achieved with 1,61% classificatory error, with using 50% of the all samples in learning and covariance range was set to 0.8. Number of created samples was 106.

In the case, tested image was “Spiral”, we changed count of samples in range from 10% to 50% of the all samples, used in classification. Sample step was set to 10. Covariance range we changed in range from 0.5 to 1.4 in step 0.15.

We found out, that the best results were achieved with 9.04% classificatory error, with using 50% of the all samples in learning and covariance range was set to 1.1. Number of created samples was 57.

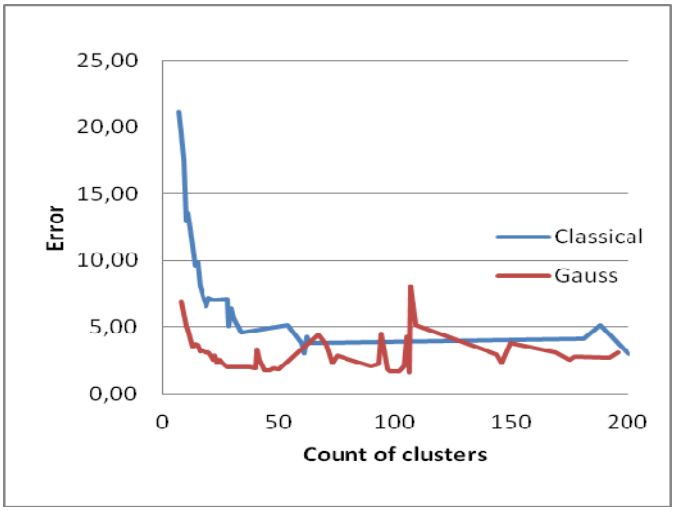


Fig. 5 Relation of classical MF ARTMAP neural network and modified MF ATRTMAP neural network for the image circle in the square.

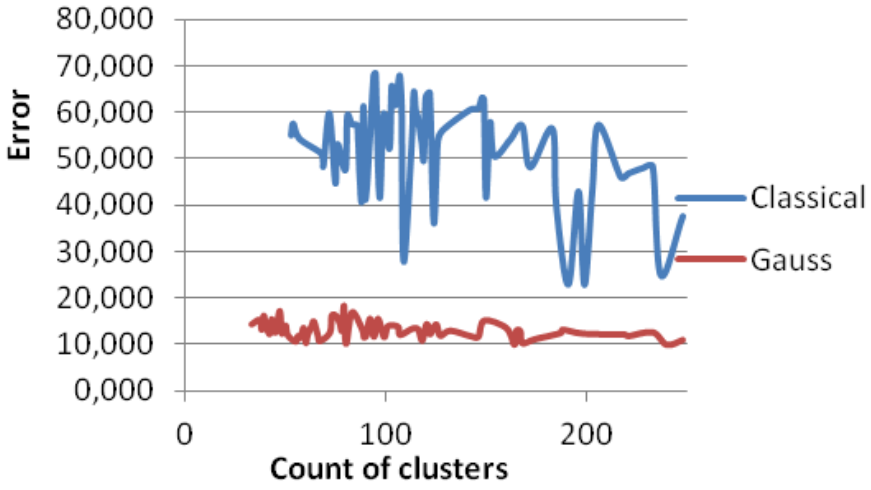


Fig. 6 Relation of classical MF ARTMAP neural network and modified MF ATRTMAP neural network for the Spiral image.

IV. Conclusion

We can see, that modified MF ARTMAP neural network has better results using smaller number of clusters. Speed of the leraning is in some cases slower and in other cases bigger, but overalll is modification better choice for classification images circle in the squrae an spriral as classical approach.

Acknowledgment

This work is the result of the project implementation: Development of the Center of Information and Communication Technologies for Knowledge Systems (ITMS project code: 26220120030) supported by the Research & Development Operational Program funded by the ERDF.

REFERENCES

[1] G. A. Carpenter, S. Grossberg, The ART of Adaptive Pattern Recognition by a Self-Organizing Neural Network, Computer, 21(3): 77-88, 1988

[2] G. A. Carpenter, S. Grossberg, and J. H. Reynolds. ARTMAP: Supervised Real-time Learning and Classification of Nonstationary Data by a Self-organizing Neural Network. Neural Networks, 4(5):565–588, 1991.

[3] R.O. Duda, P.E. Hart, and D.G. Stork. Pattern Classification, Second Edition. John Wiley & Sons, 2000.

[4] R.P.W. Duin and E. Pekalska. Open Issues in Pattern Recognition. In Computer Recognition Systems: Proceedings of 4th International Conference on Computer Recognition Systems CORES'05 (Advances in Soft Computing), pages 27–42, Secaucus, NJ, USA, 2005. Springer-Verlag New York, Inc.

[5] S. Grossberg. Adaptive Resonance Theory. Technical Report CAS/CNS-2000-024, Boston University Center for Adaptive Systems and Department of Cognitive and Neural Systems, September 2000.

[6] P. Smolár, Improving the Accuracy of Classification and the Generalization Ability of an ARTMAP Neural Network, 11*th* Scientific Conference of Young Researchers, pp. 291-293 Herľany 2011.

Methods for the Number Modification of Clusters in Modified MF ARTMAP Neural Network

Peter SMOLÁR, Peter SINČÁK

Department of Cybernetics and Artificial Intelligence, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Slovak Republic

peter.smolar@tuke.sk, peter.sincak@tuke.sk

***Abstract* — This paper deals with modified MF ARTMAP neural network, which belongs to the family of ARTMAP neural networks and has changed equation of fuzzy relation of cluster. This network combines in own structure Adaptive Resonance Theory (ART) in learning process and fuzzy logic in output. We propose the methods, which change number of the created clusters. Two methods decrease the number of clusters and one method can increase it. Modification of number of clusters is important for optimalization of computational time of the neural network in classification process.**

***Keywords* — MF ARTMAP neural network, optimalization, computational time**

I. INTRODUCTION

In the present, it is pressure to the information technology for the fast processing with large amounts of data. Therefore new algorithms are looked for handling these requirements.

Artificial intelligent (AI) methods are also the possible ways how to work with large amount of data. Neural networks, among others AI techniques, posses effectives features: the ability to adapt to dynamically changing problems, the ability to learn and also to handle uncertainty, etc.

This paper presents possibilities, how to change the number of clusters and this way to optimalize the process of classification of samples, because, when the number of clusters is less, computational time is less and number of clusters is larger, the error might be less.

This paper is organized as follows: Section II theoretically talks about the modified MF ARTMAP neural network. Section III contains description of optimalization methods. Section IV presents results of experiment using these methods.

II. NEURAL NETWORKS OF TYPE ARTMAP

Adaptive Resonance Theory (ART) was developed by Grossberg and Carpenter [1] and was biologically inspired, trying to imitate how the brain processes information –analysis of human cognitive information and stable coding in a complex input environment [2]". Simply described, the pattern matching process compares the current input with a learnt category representation. Then this process leads to a resonant state (focusing attention and triggering category learning) or to a self-regulating parallel memory search leading to a resonant state. The important feature of ART NN is that if an established category is selected, the category´s representation may be refined to incorporate new information from the current input. In case that the search ends by selecting an antecedently untrained node, the ART NN establishes a new category.

ART NN has been used in a wide spectrum of application fields, including problems of learning and classification, recognition and prediction. Different types of ART NN have been developed: the first one by Grossberg was an unsupervised learning system used for categorize binary input patterns; the next ones were extended to categorize also the analog input patterns.

Neural networks of type ARTMAP, known also as predictive ART, have supervised learning structure, with ability to self-organize arbitrary mappings from input vectors representing features such as spectral values and terrain variables, to output vectors, representing predictions

such as vegetation classes in a remote sensing application. If fuzzy ART units are used, we talk about Fuzzy ARTMAPS. It calculates the membership function of the point from the feature space to the fuzzy class. MF ARTMAP is based on the assumption that data in feature space are organized in fuzzy clusters. Grossberg in [6] it describes as a synthesis of fuzzy logic and adaptive resonance theory. Figure 1 illustrates the basic topology of the MF ARTMAP NN and in figure 2 is illustrated membership function of fuzzy relation for one cluster.

Modification, we proposed deals with equation of fuzzy relation. For increase of accuracy with learned very complex samples we changed classical equation of fuzzy membership and we used new equation, inspired from multivariate normal (Gauss) probability distribution (See equation 1).

$$A(\vec{x}) = e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)} \quad (1)$$

where Σ is covariance matrix of input samples in feature space and x is vector of values of input sample and is μ vector of average values.

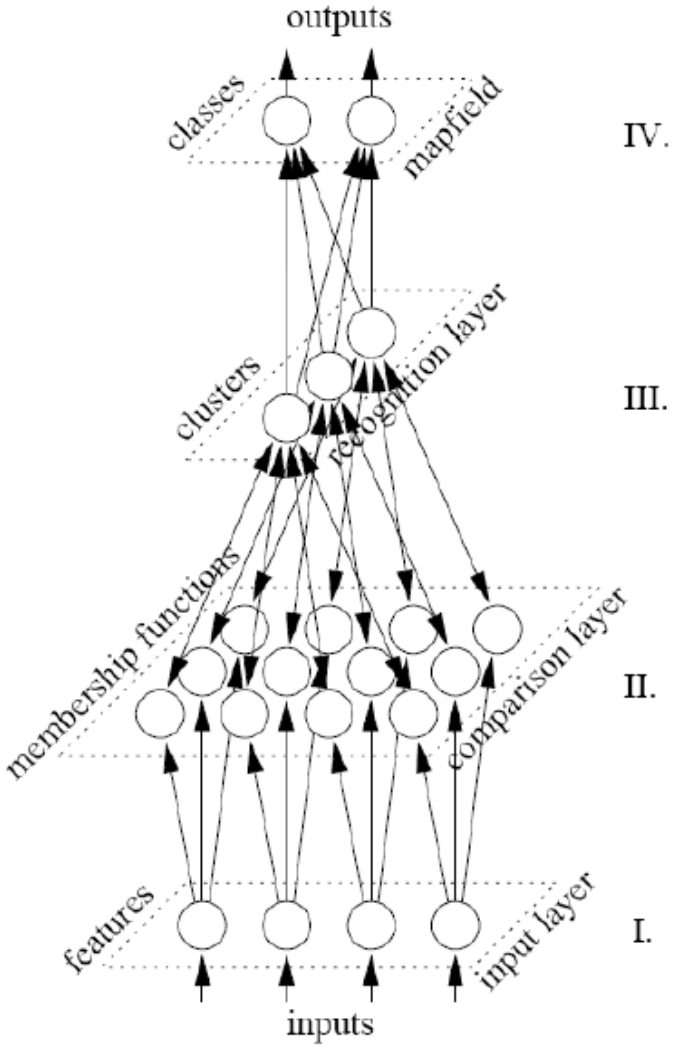


Fig. 1 Basic topology of the MF ARTMAP neural network [6].

Then clusters can better adapt to the shape of samples (Fig. 2).

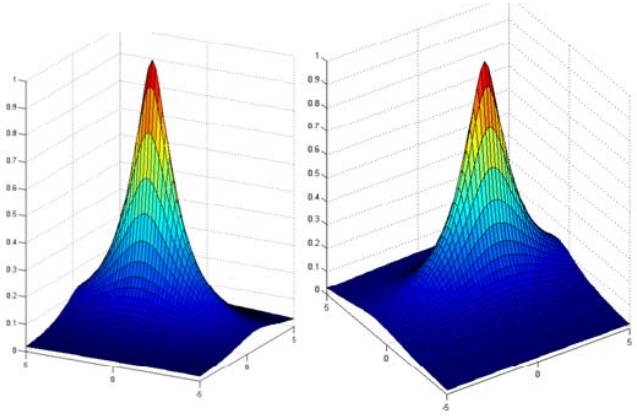


Fig. 2 Shape of the clusters.

III. METHODS FOR OPTIMALIZATION OF CLUSTER´S COUNT

As is illustrated in figure "circle within a square" (Fig. 3), if the input samples are used in a random order to MF ARTMAP network, the network generates large amounts of new fuzzy clusters depending on the parameter threshold, although these fuzzy clusters are located side by side and belong to the same fuzzy class. Ideal creation of a fuzzy clusters for an image "in the square circle" in terms of number of clusters, where the object of circle represents one fuzzy cluster.

A large number of fuzzy clusters causes that samples, which are on the edges of two clusters and belong to the same fuzzy class, have little value of membership function to the fuzzy class.

Proposed solutions to overcome this deficiency are as follows:

1. Some impact on learning has an order of training samples. Hence one of options is the selection of samples from training set to network according to the specified proximity (not exceeding the threshold parameter) to the fuzzy clusters that have already been created before. Then a new cluster is created only if no input sample can be classified to the already existing fuzzy clusters.

2. Other variety of modification is to start learning algorithms several times, but with a different sequence of input samples. Next step is unification of fuzzy relations from previous cycles of algorithm. This will increase the accuracy value of membership function of input sample to class.

3. Creation of next layer of learning algorithm in MF ARTMAP network, which will detect in the end of learning process, whether surrounding fuzzy cluster belong to the common fuzzy class or not. Then, based on fuzzy cluster environment, this cluster will "absorb" the surrounding clusters, if these fuzzy clusters belong to a same fuzzy class.

Lower number of fuzzy clusters cause faster running of algorithm and more accurate determination of membership of the sample to the class.

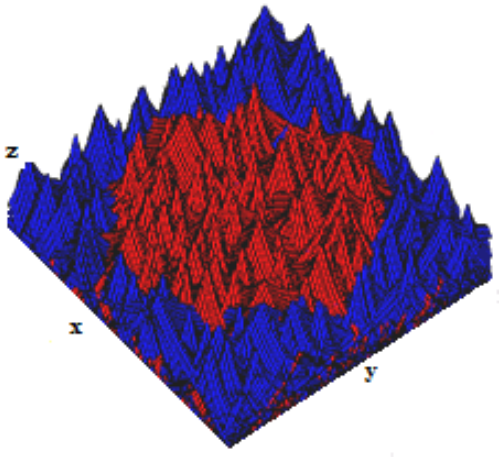


Fig. 3 Figure of large amount of created clusters using MF ARTMAP neural network

IV. Experiments with Methods for Optimalization of Cluster´s Count

We performed experiments with MF ARTMAP neural network, where we tested methods described in previous chapter.

These comparisons we did with image "circle in the square" (Fig. 4).

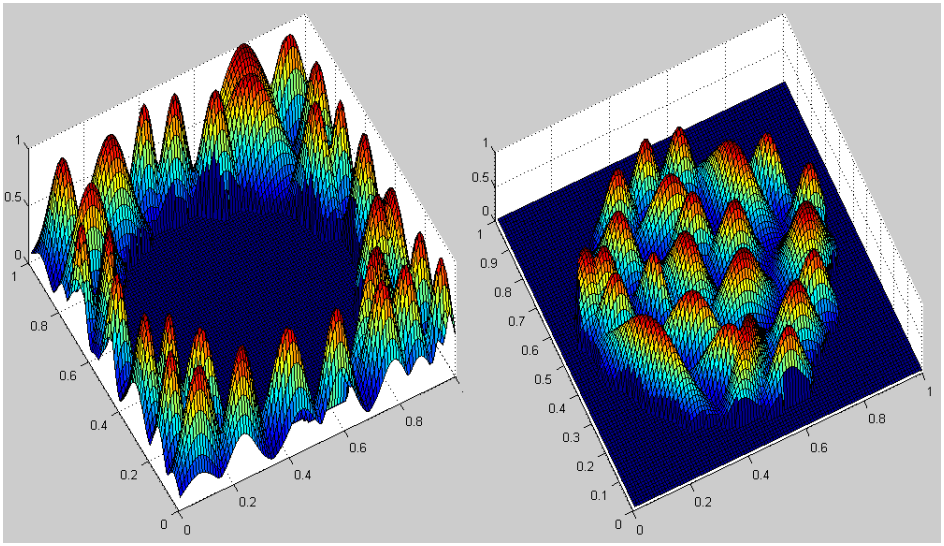


Fig. 4 Tested image, without use of methods for optimalization of clusters.

We performed experiments with all three types of methods in mathematical program MATLAB.

A. Ordering the samples before learning - preprocessing

Final clusters count was stabilized to 12. The generalization ability of the network was 0.999. Error of classification was 11.5300.

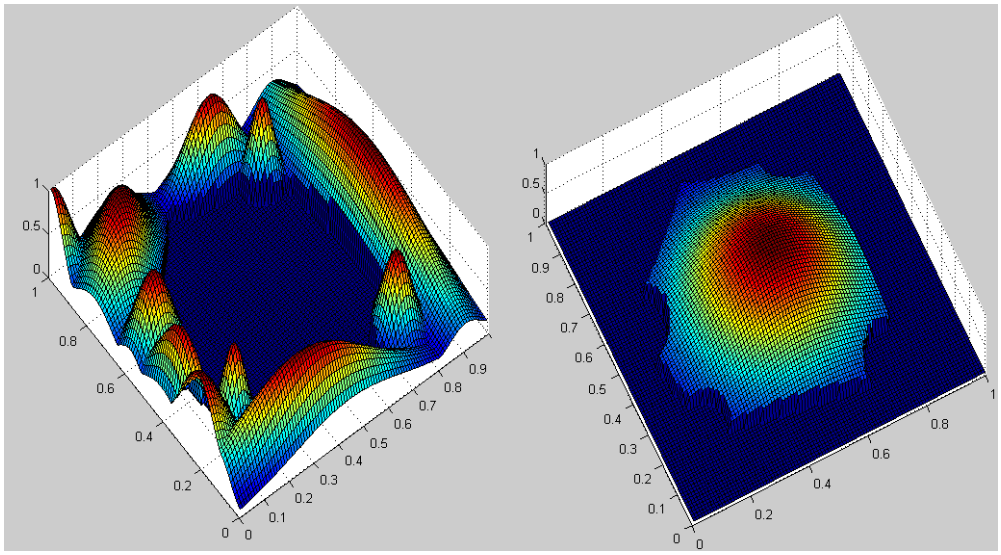


Fig. 5 Figure of optimalized number of clusters using prepreocessing (reordering) of samples.

Merging the clusters

Final count of clusters was set to 12. The generalization ability of the network was 0.999. Error of classification was 3.3000.

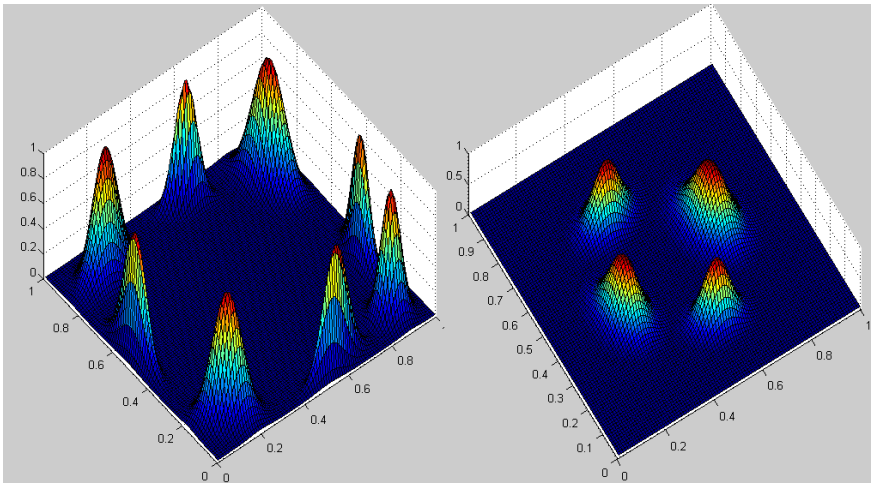


Fig. 6 Figure of merged clusters to number 12.

B. Repeating of learning

Number of learning cycles was set to 4. Entire number of cluster was achieved to 161 (In the order of cycles the new clusters were created in numbers 46, 39, 38, 41). The generalization ability of the network was 0.9836. Error of classification was 2.5500.

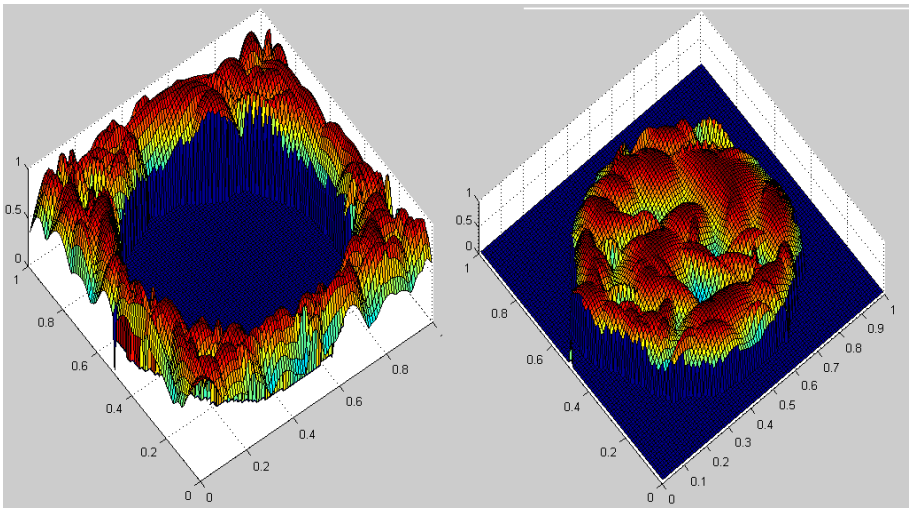


Fig. 7Figure of all clusters after running of learning process.

V. CONCLUSION

We found out, that all three methods have some advantages and also some disadvantages. Advantages of first two methods are, the computational time was less, but error gently increased. Third method can increase accuracy, but computational time markedly increased. But they are suitable methods for controlled manipulation with number of clusters.

ACKNOWLEDGMENT

This work is the result of the project implementation: Development of the Center of Information and Communication Technologies for Knowledge Systems (ITMS project code: 26220120030) supported by the Research & Development Operational Program funded by the ERDF.

REFERENCES

[1] G. A. Carpenter, S. Grossberg, The ART of Adaptive Pattern Recognition by a Self-Organizing Neural Network, Computer, 21(3): 77-88, 1988

[2] G. A. Carpenter, S. Grossberg, and J. H. Reynolds. ARTMAP: Supervised Real-time Learning and Classification of Nonstationary Data by a Self-organizing Neural Network. Neural Networks, 4(5):565–588, 1991.

[3] R.O. Duda, P.E. Hart, and D.G. Stork. Pattern Classification, Second Edition. John Wiley & Sons, 2000.

[4] R.P.W. Duin and E. Pekalska. Open Issues in Pattern Recognition. In Computer Recognition Systems: Proceedings of 4th International Conference on Computer Recognition Systems CORES'05 (Advances in Soft Computing), pages 27–42, Secaucus, NJ, USA, 2005. Springer-Verlag New York, Inc.

[5] S. Grossberg. Adaptive Resonance Theory. Technical Report CAS/CNS-2000-024, Boston University Center for Adaptive Systems and Department of Cognitive and Neural Systems, September 2000.

[6] A. K. Jain, M. N. Murty, and P. J. Flynn. Data Clustering: A Review. ACM Comput. Surv., 31(3):264–323, 1999.

[7] P. Smolár, Improving the Accuracy of Classification and the Generalization Ability of an ARTMAP Neural Network, 11th Scientific Conference of Young Researchers, pp. 291-293 Herľany 2011.

Opinion classification based on dictionary generation

Peter SZABÓ, Kristína MACHOVÁ

Department of Cybernetics and Artificial Intelligence, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Slovak Republic

peter.szabo.2@tuke.sk, kristina.machova@tuke.sk

***Abstract* — This paper focuses on the area of sentiment analysis. It specifies the importance of opinion in the context of automatic sentiment and opinion analysis. It also divides the applications of sentiment analysis related to social web into different groups. The article proposes the algorithm for dictionary generation and provides conclusion in conjunction with experiments.**

***Keywords* — dictonary approach, dictionary generation, exogenous methods, sentiment analysis, sentiment classification**

I. INTRODUCTION

Language, as an ability of human communication contains a large amount of expressions and expressive possibilities. In general, the information which is formulated in some language can be the fact or the opinion [1]. The fact is objective information about an entity, event or theirs properties. On the other hand the opinion usually reflects a subjective approach which describes the human feelings and evaluates the entities, events and properties. An extensive research has been carried out in the field of knowledge processing and discovery (data mining, text classification) but substantially less work has been done in the field of opinion mining. And just the opinion is the key element that affects human decisions and human behavior.

The importance of the automatic sentiment analysis and its subsequent evaluation is indisputable. This approach was applied on the social web within various applications [2] [3] [4] [5]. These applications were mainly focused on the following fields:

1. product and service benchmarking (informative feedback about quality of various products and services)
2. market intelligence (acquisition and accumulation of market information from publicly available sources)
3. advertisements placement (targeted advertising based on the addressee's opinion of the product)
4. opinion retrieval (general search for opinions)

II. OPINION ANALYSIS WITHIN SOCIAL WEB

According to [5], there are more than 20 companies engaged in the analysis of sentiment and offering services based on opinion mining and opinion classification within the United States. The following sections introduce various areas, which sentiment and opinion analysis is used in.

A. Sentiment and subjectivity classification

Sentiment and subjectivity analysis is a domain where the problem of sentiment and opinion analysis is considered as a classification problem. There are two different approaches to opinion classification. At first, classification of the whole block of text (for example document, blog, etc.) to positive or negative class. At second, classification of individual sentences to subjective and objective classes. The polarity of a sentence with subjectivity can be positive, neutral or negative. While the first approach to the opinion classification is focused on some author's opinion (for example opinion on product, a political decision, etc.), the second approach is focused on identification of those sentences that carry feelings and attitudes.

Having found such sentences, it determines their polarity [2].

B. Feature based sentiment analysis

Feature based sentiment analysis searches for an entity composed of given certain parts, properties and then determines its polarity. The entity may be a product, a service, a person, an organization, an event, etc. The fundamental difference to the previous method is the possibility to identify some object and its parameters and to determine the polarity of a selected parameter or property of the object (e.g., mobile and its display).

C. Sentiment analysis based on comparison

Sentiment analysis of comparative sentences focuses on comparison of two entities and determines which entity is better than another. Within the previous fields the opinion was obtained directly and there were no comparison to other entities. This access to sentiment analysis is based on the search for the text which is based on confrontation and creates the conclusion (mobile phone A is better than mobile phone B).

D. Opinion search

Opinion search and retrieval is an area where opinions are searched on the basis of a given set of keywords, which express opinions, the user is interesting in. In this way, it is possible to find positive and negative opinions about the specified entity (Nokia E51) [6].

E. Opinion spam

Opinion spam is an area where text search is performed within discussion forums with the purpose of finding those opinions that are not expected and/or useful for the discussed issue. In this area, we can talk about classification problem because the searched text is classified to one of two classes: spam or non–spam class.

F. Opinions utility determination

The last research area is utility of opinions which corresponds to the usefulness and quality of opinions. Rated opinions can be sorted and the reader is able to obtain opinions of the highest quality.

III. Particular language processing by dictionary approach

In general, the methods that are used in sentiment analysis can be inserted to exogenous and endogenous approaches. Endogenous methods of opinion analysis use methods of machine learning and training examples for estimation and prediction of the polarity of words. Exogenous methods extend an endogenous method by an external source of knowledge which is usually a dictionary. Most exogenous methods are based on the analysis and particular processing of natural language.

Interesting possibility of exogenous methods is a dictionary generation automatically or semi automatically from discussions texts. This approach can avoid the need to annotate texts manually by people. The Slovak language obtains variety of expressing possibilities and every word can gain different shapes, forms and can be placed in different parts of the sentence, which can change its meaning. Dictionary generation for Slovak language can bring more advantages than its generation for example for English.

IV. Dictionary generation

The grammatical meaning in Slovak language is usually expressed by different form of the same word. For example:

- I am going, “ide-m”
- You are going, “ide-š”
- We are going, “ide-me”

This is called declension. Thanks to declension and set of words (in different forms), it is possible to classify the word to one of the lexical category and specify its grammatical meaning.

The application has been made to identify a nouns (and its declension model) from input text, based on different form of word endings. The nouns have been classified to 14 declension models with certain probability.

The probability is determined using the following algorithm:

1. search for the words with common root (“Zuzka”, “Zuzky”, “Zuzkám”)

2. extract word endings (“a”, “y”, “ám”) and look up for the endings in each of 14 declension models
3. for each declension model calculate: (number of same different endings / total different endings for current model) * 100

V. EXPERIMENTS

The first experiment has been done using the following input data: “Dostal som darčeky. V prvom darčeku bolo autíčko. K ďalším darčekom som sa nedostal.”

The application has evaluated the input and 12 (two models have been evaluated to zero) declension models has been assigned with the following probabilities:

Table 1
The declension model probabilities for the noun "darček".

Declension model	Probability
Chlap	*25.00%*
Hrdina	*25.00%*
Dub	***44.44%***
Stroj	*33.33%*
Kuli	*22.22%*
Žena	*33.33%*
Ulica	*22.22%*
Dlaň	*12.50%*
Kosť	*14.29%*
Gazdiná	*11.11%*
Mesto	*30.00%*
Srdce	*30.00%*

The application has correctly identified the noun “darček” with right declension model “dub”.

The second experiment has been done using the following input data: “Zuzka je pekné dievča. Zuzkám som sa vždy páčil. Každý deň o Zuzke snívam.”. The noun “Zuzka” and declension model “žena” (with the probability 44%) has been correctly identified:

Table 2
The declension model probabilities for the noun "Zuzka".

Declension model	Probability
Chlap	*25.00%*
Hrdina	*12.50%*
Dub	*33.33%*
Stroj	*33.33%*
Kuli	*11.11%*
Žena	***44.44%***
Ulica	*33.33%*
Dlaň	*25.00%*
Kosť	*14.29%*
Gazdiná	*22.22%*
Mesto	*40.00%*
Srdce	*30.00%*
Dievča	*10.00%*

The third experiment has been done using the text from the folk tale “Rozprávka o leňochii” with exactly 589 words. An algorithm has correctly identifies 19 nouns with its declension models.

VI. Conclusion

Natural language processing is not a simple task and the complexity of solving problems related to the sentiment analysis is increasing [2].

The application proposed in this paper has several disadvantages:

- The algorithm is not robust against keying mistakes - if the word is incorrect, the evaluation will be incorrect as well.
- It is necessary to deliver correct grammar for the given input.
- Slovak language contains large number of grammatical exceptions. It is not possible to identify this words using proposed algorithm.
- Input data have to contain sufficient amount of words in different forms.

An identification of the other lexical categories will be object of the next work.

Acknowledgment

This work is the result of the project implementation Development of the Centre of Information and Communication Technologies for Knowledge Systems (project number: 26220120030) supported by the Research & Development Operational Program funded by the ERDF (100%).

References

[1] B. Liu, "Sentiment analysis and opinion mining", http://www.cs.uic.edu/~liub/FBS/Sentiment-Analysis-tutorial-AAAI-2011.pdf.

[2] A. Beal, "Is Automated Sentiment Analysis Reliable?", http://www.marketingpilgrim.com/2009/08/why-sentiment-analysis-is-about-as-reliable-as-a-canary-in-a-coal-mine.html

[3] P. Yared, "Why sentiment analysis is the future of ad optimization", http://venturebeat.com/2011/03/20/why-sentiment-analysis-is-the-future-of-ad-optimization/

[4] M. Russell, "With sentiment analysis, context always matters", http://radar.oreilly.com/2011/03/sentiment-analysis-context.html

[5] B. Liu, "Sentiment Analysis and Subjectivity", http://www.cs.uic.edu/~liub/FBS/NLP-handbook-sentiment-analysis.pdf

[6] K. Eguchi, V. Lavrenko, "Sentiment retrieval using generative models.", Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 345–354 (2006)

Prečo monitorovať on-line komunikáciu?

Ján ŠTOFA, Kristína MACHOVÁ

Katedra kybernetiky a umelej inteligencie, Fakulta elektrotechniky a informatiky,
Technická univerzita v Košiciach, Slovenská republika

jan.stofa@tuke.sk, kristina.machova@tuke.sk

***Abstrakt* — Predkladaný príspevok sa sústreďuje na oblasť bezpečnosti v rámci sociálneho webu. Popisuje komunikáciu prostredníctvom internetu a hrozby s tým súvisiace. V rámci príspevku je rozvíjaná problematika kontroly detskej web komunikácie. V princípe ide o nahrávanie slov priamo z konkrétnej on-line komunikácie, ktorá sa spracúva tak, aby slovník obsahoval, čo najvýstižnejšie jazyk aktérov danej komunikácie (dieťaťa – obete a manipulátora - ohrozovateľa). Problematika bezpečnosti je reálnou súčasťou sociálnych sietí. Príspevok ponúka nový prístup k ďalšiemu smerovaniu výskumu v danej oblasti využitím klasifikácie názorov a dolovania v dátach na generovanie charakteristických slov webovej komunikácie. Následne sú tieto webové príspevky klasifikované do dvoch tried: znamená hrozbu a neznamená hrozbu pomocou slovníka generovaného za účelom riešenia daného problému.**

***Kľúčové slová* — bezpečnosť, sociálna sieť, webová komunikácia, klasifikácia názorov, slovníkový prístup**

I. Úvod

Komunikácia je základným prvkom medziľudskej interakcie. Komunikovaním si hovoríme významy v priamom alebo nepriamom kontakte a je to proces dorozumievania sa medzi ľuďmi. Je považovaná za základný spoločenský proces.

Dnes je už bežná komunikácia, pri ktorej nehľadíme do tváre nášho partnera, ale na monitor nášho osobného počítača. Vplyv internetu ako komunikačného média dnešnej doby je nepopierateľný, ale neprináša so sebou iba pozitíva a výhody, ale aj riziká a problémy, ktoré sa priamo týkajú nášho osobného života, či životného štýlu.

Jedným z hlavných prínosov internetovej komunikácie je umožnenie informačnej dostupnosti, interaktívnosti pri výmene správ a rýchlosti pri sprostredkovaní kontaktu. Na druhej strane môže mať využívanie internetu dopad nielen na kvalitu reči a myslenia, ale aj na celkový osobný život danej osobnosti.

Off-line komunikácia, alebo neinteraktívna komunikácia je najčastejšie prezentovaná e-mailom a diskusnými fórami, zatiaľ čo on-line komunikácia predpokladá odozvu v reálnom čase a vyžaduje súčasnú prítomnosť aspoň 2 účastníkov procesu. Tento typ komunikácie je medzi deťmi, dospievajúcimi, ale aj dospelými veľmi obľúbený, najmä vďaka anonymite, vizuálnej atraktívnosti, nezávislosti na čase a priestore...

Užívatelia virtuálnej komunikácie si môžu svoju komunikáciu premyslieť, môžu komunikovať príjemné, pozitívne a dôležité informácie. V komunikácii však môžu aj urážať, napádať, vydierať, vyhrážať sa a manipulovať, nútiť používateľov rôznymi spôsobmi k osobnému kontaktu...

Jedným z nebezpečenstiev pre bežných používateľov pri on-line komunikácii môže byť aj anonymita. Tá poskytuje priestor pre rôzne typy nebezpečných komunikačných techník (napr. phishing, pharming, grooming). Tieto spôsoby komunikácie so sebou prinášajú problémy, ktoré hraničia s nezákonným správaním a môžu viesť k psychickému a fyzickému poškodeniu osobnosti.

Prostredníkom pre on-line komunikáciu sú sociálne siete. Sú to komunikačné kanály, ktoré slúžia ako galérie a adresáre kontaktov. Ide o internetové aplikácie – programy (softvér), ktoré si vyžadujú špecifický prístup. Sú charakteristické predovšetkým týmito vlastnosťami:

- predstavujú priestor pre konverzáciu
- poskytujú možnosť zdieľania obsahu

- ponúkajú priestor pre publikovanie obsahu tvoreného ich používateľmi
- podporujú vzájomnú participáciu používateľov na tvorbe obsahu.

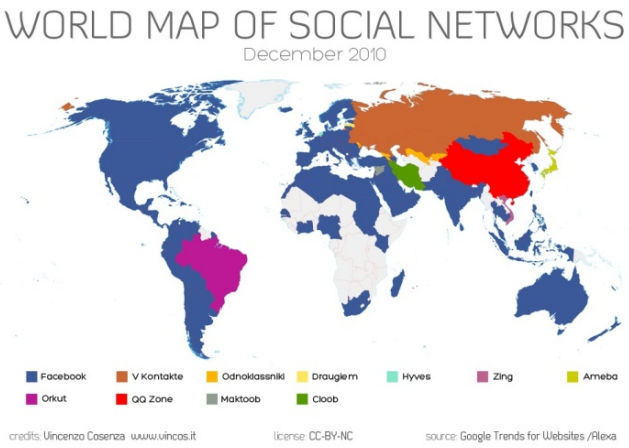
Kombinácia týchto vlastností dáva priestor aktivitám a funkciám ako je budovanie vzťahov, distribúcia obsahu tvoreného používateľmi, hodnotenie a značenie obsahu a samozrejme zábave. V rámci užívateľského obsahu nejde iba o textový obsah, ale aj o fotografie a videá.

Medzi najznámejšie sociálne siete používané slovenskými používateľmi patrí predovšetkým Facebook, slovenský Pokec, MySpace, Twitter, LinkedIn, GoogleBuzz, či YouTube [1].

II. PREHĽAD PROBLEMATIKY PROSTREDNÍCTVOM PRIESKUMOV

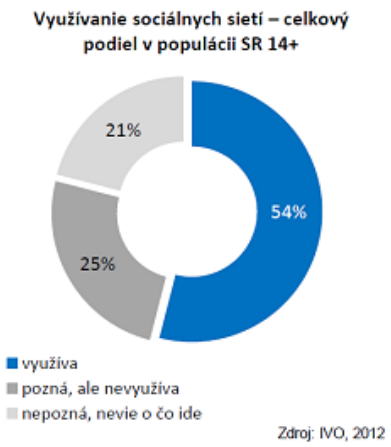
O tom, ako sociálne siete zaplavujú svet a pomaly začínajú ovplyvňovať a niekedy až ovládať životy ľudí svedčia aj nasledujúce zistenia. Ide o rôzne svetové i naše prieskumy uskutočňované s cieľom zmapovať používanie sociálnych sietí. Za zmienku stoja aj prieskumy, ktoré zisťujú najnižšiu vekovú hranicu ich používateľov.

Mapa sveta podľa sociálnych sietí zobrazená na obrázku č.1 ruší zdanie, že svetom vládne Facebook. Aj keď je u nás a vo väčšej časti Európy fenoménom práve Facebook, krajiny ako je Brazília Rusko, Afrika a Čína tomuto trendu zatiaľ nepodľahli. Pri pohľade na túto mapu však naďalej ostáva faktom, že Facebook so svojimi 600 miliónmi používateľov je vo svete dominantnou sociálnou sieťou. Zo 132 analyzovaných krajín sa rozmohol v 115 krajinách sveta. Ďalšie siete, ktoré nasledujú za Facebookom sú napr. Twitter, či Linkedln (Austrália, Kanada, Anglicko, Taliansko, Nemecko).



Obr. 1 Svetová mapa sociálnych sietí, zdroj: [2]

Jedným z prieskumníkov bol aj Inštitút pre verejné otázky (IVO). Zrealizoval prieskum digitálnej gramotnosti a informatizácie spoločnosti. Zameriaval sa najmä na zmapovanie podielu a intenzity využívania sociálnych sietí vo všeobecnosti, ako aj na preferenciu niektorých typov sietí a účely, na ktoré sú tieto siete najčastejšie využívané. Zobrazený diagram poukazuje na celkový podiel využívania sociálnych sietí v populácii SR nad 14 rokov.

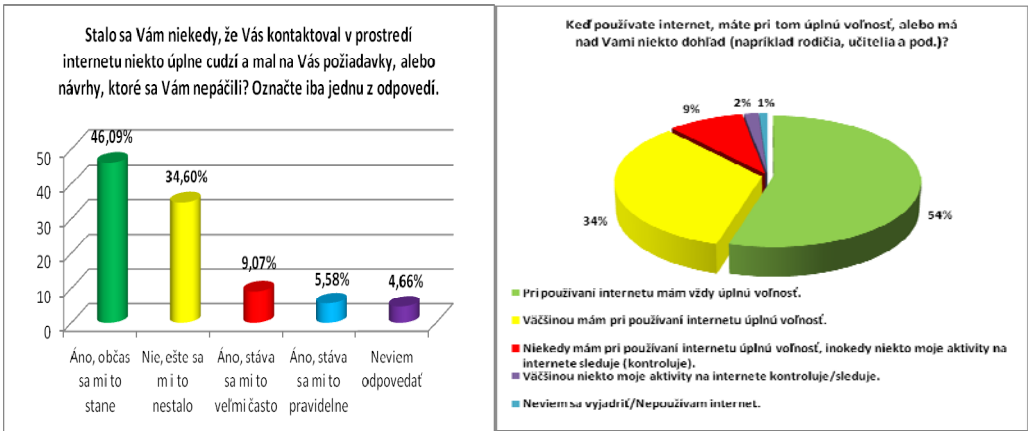


Obr. 2 Výsledky výskumu IVO 2012, zdroj: **Chyba! Nenašiel sa žiaden zdroj odkazov.**

Okrem iného sa výskum zameral aj na vekové zastúpenie populácie pri používaní sociálnych sietí. V populácii nad 14 rokov využíva sociálne siete až 54 % respondentov. Štvrtina populácie síce sociálne siete pozná, ale z rôznych dôvodov ich zatiaľ nevyužíva. Zvyšných 21 % populácie sociálne siete vôbec nepozná.

Sociálne siete sú nepochybne skvelým komunikačným prostriedkom, no tento virtuálny svet má aj svoje temné stránky. Príkladom je, keď si kriminálnici nájdu svoje obete práve na jednej zo sociálnych sietí.

Významný prieskum uskutočnilo aj Ministerstvo školstva SR v roku 2009. Výsledky výskumu sú síce z časového hľadiska dnes už neaktuálne, ale s najvyššou pravdepodobnosťou ide v súčasnosti o stúpajúcu tendenciu týchto čísel, ktoré majú negatívny dopad práve na deti a mládež. Z množstva položených otázok vyberáme len dve, ktoré priamo poukazujú na akútnosť a aktuálnosť riešenej problematiky.



Obr. 3 Výsledky výskumu MŠ SR 2009, zdroj: [4]

III. PREČO MONITOROVAŤ SOCIÁLNE SIETE?

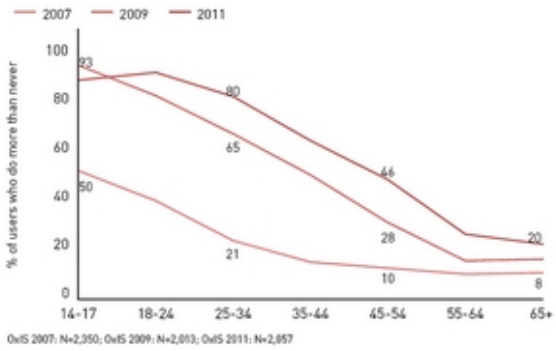
Na monitoring sociálnych sietí a on-line komunikácií možno nazerať z dvoch hľadísk. Prvé hľadisko má na zreteli pozitívne využitie sociálnych sietí vo svoj prospech. Máme na mysli firemný marketing a personálny manažment pri výbere potenciálnych zamestnancov. Mnohé firmy využívajú sledovanie diskusií o svojich produktoch a značke v diskusnom fóre na niektorých zo sociálnych sietí. Monitoring tak predstavuje efektívny nástroj, ktorý firmám pomáha urobiť si okamžitú predstavu o tom, čo ľudí práve zaujíma. Firma tak dokáže pohotovo reagovať na potreby zákazníkov. Získava rýchlu spätnú väzbu, ako aj informácie o konkurencii. Problematikou získavania informácií zo sociálnych sietí sa zaoberá klasifikácia názorov (opinion mining, sentiment classification, sentiment analysis). Ide o účinný spôsob, ktorého cieľom je zistiť postoj jednotlivého prispievateľa, resp. webovej diskusie ako celku, k určitej téme. Významné postavenie má názor, čo je pohľad, postoj alebo hodnotenie objektu držiteľom názoru. Pre takýto marketingový monitoring on-line komunikácie má väčšiu váhu konkrétny názor, postoj ako konkrétny autor názoru. Pri klasifikácii názorov je dôležité, že nie všetky slová textu sú rovnako významné pri odhaľovaní a klasifikácii názoru. Nositeľmi subjektívnych postojov v texte sú hlavne prídavné mená, príslovky, podstatné mená a slovesá. Je preto nutné identifikovať v texte slová, ktoré majú význam pri klasifikovaní názorov. Problém, s ktorým sa pri klasifikácii názorov stretávame je určenie subjektivity, orientácie (polarity) slov a určenie sily orientácie [5].

Druhý pohľad predstavuje monitoring sociálnych sietí z hľadiska globálnej bezpečnosti. Touto problematikou sa zaoberá aj legislatíva a konkrétne zákony SOPA, PIPA a CISPA. Aj keď ide o premyslenú kontrolu a monitorovanie týchto sietí z hľadiska kybernetickej bezpečnosti, sú tieto legislatívne návrhy pod paľbou kritiky hlavne zo strany širokej verejnosti. Mnoho ľudí sa obáva, že bude narušené ich právo na súkromie a celá táto kontrola sa zvrhne na prísnu cenzúru virtuálneho sveta. Cieľom právneho predpisu H. R. 3523 – CISPA (Cyber Intelligence Sharing and Protection Act), ktorý bol len nedávno vytvorený, však má byť účinný boj proti kybernetickým útokom [6]. Podľa nás je tento spôsob legálnej kontroly on-line komunikácie prijateľný.

Otvára sa tak možnosť sledovať komunikáciu medzi aktérmi sociálneho webu a konkrétne sa

zamerať na vopred určené kľúčové slová, znamenajúce hrozbu, ako aj na frekvenciu ich výskytu v bežnej sieťovej komunikácii. Týmto spôsobom by sa dala zvýšiť bezpečnosť nielen jednotlivých používateľov rozličných webových portálov, ale napomohlo by to aj k zvýšeniu bezpečnosti ľudí vo všeobecnosti. Pojem bezpečnosť je veľmi všeobecný a zahŕňa v sebe veľké množstvo sledovaných oblastí. Tie predstavujú aj široké spektrum možnosti, na ktorých by sa dal výskum realizovať. Takýto výskum by znamenal možnosť predchádzať napríklad teroristickým útokom a inému násiliu. Monitoringu by tak boli vystavení aktéri komunikácie na sociálnych sieťach, ktorí sa na nich pohybujú s cieľom upútať pozornosť, získať obeť, prezentovať svoje radikálne náboženské či politické názory, vyvolať paniku.

Tieto témy v sebe obsahujú mnoho rozdielnych slov, ktoré je potrebné sledovať a porovnávať. Toto porovnávanie je nutné realizovať s vopred vytvoreným slovníkom. Pretože pre účely klasifikácie názorov je potrebné automaticky vytvoriť slovník termov, ktoré sú nositeľmi názoru v danej oblasti. Keďže by bolo veľmi náročné vytvoriť slovník aktuálny pre globálnu bezpečnosť ako takú, je nutné vybrať si jednu konkrétnu oblasť. Preto sme sa rozhodli podrobnejšie sa venovať dnes veľmi aktuálnej oblasti ako je zneužívanie detí prostredníctvom sociálnych sietí. Na rizikovú skupinu detí a mládeže ohrozovanú v kyberpriestore poukazuje aj nižšie zobrazený graf.



Obr. 4 Percento používateľov sociálnych sietí podľa vekových kategórií, zdroj: [7]

Z grafu vývoja používania sociálnych sietí vyplýva niekoľko kľúčových informácií. Medzi rokmi 2007 a 2009 bol zaznamenaný dramatický nárast počtu používateľov sociálnych sietí. Tento nárast je spôsobený aj extrémnou popularitou sociálnych sietí medzi tínedžermi. Graf síce zachytáva vekovú skupinu mladistvých od štrnásť rokov, no v skutočnosti sa v priestore sociálnych sietí pohybuje veľké percento maloletých pod štrnásť rokov.

Spomínaný slovník termov zameraný na oblasť zneužívania detí sa javí byť pomerne jasný a zrozumiteľný, pretože útočníci používajú zaužívané a jednoducho rozpoznateľné metódy. Pomocou nich si získavajú svoje obete a následne s nimi udržiavajú kontakt. Takto nami vytvorený slovník by obsahoval slová a slovné spojenia typické pre tento druh manipulácie s deťmi. Základným manipulačným prostriedkom týchto „kyberkriminálnikov" je presvedčivo využiť jazyk, ktorý umožňuje získať potrebné údaje o obeti a zmanipulovať ju tak, aby prezradila čo naviac informácií o sebe. Tak sa manipulátor jednoduchšie dostane k svojej obeti. V takejto komunikácii možno zvýrazniť niekoľko etáp, ktoré medzi sebou manipulátor kombinuje, aby dosiahol svoj zámer:

Etapa izolácie obete od okolia:

- vzbudiť dôveru a izolovať obeť od okolia
- manipulátor je v pozícií chápajúceho a riešiteľa problému obete
- manipulátor vytvára dojem najlepšieho priateľa
- používa frázy:
    - Rodičia ti nerozumejú, ja áno.
    - Mne sa môžeš zdôveriť so svojimi problémami.
    - Nehovor o tom rodičom.
    - Keby si im to povedal, budú nám brániť rozprávať sa (písať si).
    - Dospelí to nepochopia, ja áno.
    - Som na tom podobne, zver sa mi.

Etapa získavania základných informácií o manipulovanom:

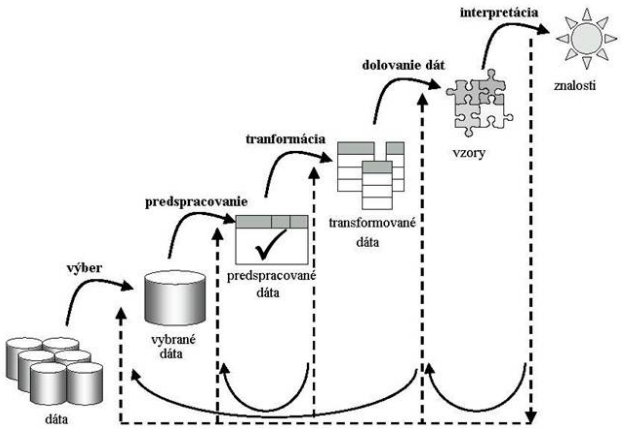
- po izolácii dieťaťa manipulátor zisťuje základné informácie o obeti

- manipuluje dieťa prostredníctvom darčekov (napr. PC hra)
- pokus o získavanie mobilného čísla obete a ďalších osobných informácií
- používa otázky:
  - Kde bývaš? Ja v Košiciach.
  - Čo rád robievaš? Ja mám rád počítačové hry.
  - Koľko máš rokov? Ja pätnásť.
  - Pošli mi fotku, ja ti pošlem svoju.

Etapa priamej manipulácie (vydieranie):

- jej cieľom je poškodiť obeť – snaha dosiahnuť osobnú schôdzku, či donútiť k videohovoru
- vydieranie je podmienené intenzitou virtuálneho vzťahu obete a manipulátora
- časté frázy:
  - Ak sa so mnou nestretneš, zabijem sa.
  - Potrebujem s Tebou hovoriť osobne [8].

Bezpečnosť detí na sociálnych sieťach vnímame ako problematiku, ktorú je možné riešiť v spojitosti s klasifikáciou názorov. Na rozdiel od klasifikácie názorov pri monitoringu firemného marketingu, kde sa napríklad v rámci sily orientácie slova orientovalo na negativitu a pozitivitu názorov, v prípade monitoringu detskej on-line komunikácie plánujeme vytvoriť slovník orientovaný na vyššie spomenuté frázy a jednotlivé slová. Konkrétnym termom bude pridelený stupeň možného ohrozenia. Keďže sa v našom výskume chceme zamerať na detské chaty a skupiny budeme postupovať na báze procesu objavovania znalostí (Knowledge Discovery in Databases - KDD).



Obr. 5 Proces objavovania znalostí, zdroj: [9]

Proces objavovania znalostí (obr. č.5) je interaktívny a iteratívny, pričom zahŕňa viaceré kroky s množstvom rozhodnutí vykonaných používateľom. Vo všeobecnosti možno povedať, že tento proces pozostáva z nasledovných základných krokov (znázornené na obrázku):

Pochopenie aplikačnej domény – riešiteľ musí pochopiť a identifikovať kľúčové koncepty aplikačnej domény, ktorou je problematika bezpečnosti a stanoviť cieľ procesu KDD.

Vytvorenie cieľovej množiny dát – získavanie relevantnej množiny dokumentov zo sociálnych sietí (detské chaty a diskusné skupiny). Vo vybraných dátach sa budú ďalej hľadať znalosti.

Predspracovanie dát – zahŕňa rôzne techniky spracovania prirodzeného jazyka (lexikálna, syntaktická a sémantická analýza). V tomto procese dochádza k redukcii (odstránenie neplnovýznamových slov, prevedenie rôznych tvarov toho istého slova na kmeň) a následne k uloženiu množiny dokumentov v internej forme reprezentácie.

Dolovanie v textoch – ide o samotnú aplikáciu zvoleného algoritmu dolovania v textoch, klasifikácia.

Vyhodnotenie a interpretácia výsledkov – vizualizáciu poskytujú technické nástroje na vhodnú reprezentáciu nájdených vzorov. Samotné vyhodnotenie a interpretácia výsledkov sú závislé od riešiteľa [10].

Ako veľmi účinná metóda pre riešenie uvedenej problematiky je nahrávanie slov do klasifikačných slovníkov. V princípe ide o nahrávanie slov priamo z konkrétnej on-line komunikácie, ktorá sa spracúva tak, aby slovník obsahoval čo najvýstižnejšie jazyk aktérov

danej komunikácie (dieťaťa – obete a manipulátora - ohrozovateľa). Bohatosť slovníka znamená presnejšiu klasifikáciu názorov.

Výstupom tohto výskumu sa predpokladá aplikácia určená na detekciu hrozieb vyplývajúcich z on-line komunikácie na detských chatoch a v diskusných skupinách. Takáto aplikácia by mala pomôcť pri identifikácii hrozbu predstavujúcej komunikácie. Pri testovaní tejto aplikácie budú použité diskusné príspevky zo sociálnych sietí.

IV. ZÁVER

Na základe zhody takto vytvoreného slovníka termov s príspevkami vo webových diskusiách by bolo možné identifikovať komunikáciu s podozrivou povahou a pravdepodobne aj jej autora. Našou snahou bude vytvoriť taký systém, ktorý bude sám vyhodnocovať danú on-line komunikáciu, a to tak, že v nej vyhľadá a zistí frekvenciu výskytu slov nami označených ako podozrivých. Ak táto frekvencia prekročí vopred definovanú prahovú hodnotu, daná komunikácia bude vyhodnotená a označená ako podozrivá. Tento proces považujeme za dôležitý prvok pri riešení tak závažného problému súčasného sveta ako je zneužívanie detí a mládeže prostredníctvom sociálnych sietí, či všeobecne vo virtuálnom svete internetu. Tento nový fenomén kyberpriestoru sa nazýva grooming (on-line zneužívanie). Táto problematika však úzko súvisí s ďalšími nebezpečenstvami, ktoré číhajú na maloletých a mladistvých prostredníctvom webu. Najnovšie sa ukazuje byť postupne sa rozširujúcim a závažným problémom kyberšikana. Tá sa už ale netýka iba vzťahu dieťa - dospelý, ale aj vzťahov dieťa - dieťa a dospelý – dospelý. To je jeden z ďalších podnetov pre nás a náš výskum, aby sme spomínanú aplikáciu zrealizovali a pracovali na jej zdokonaľovaní, aby sa stala skutočným pomocníkom najmä pre rodičov chrániacich svoje ratolesti, no aj pre zamestnávateľov, ktorý vďaka tomuto prístupu budú môcť včas odhaliť kyberšikanu, či mobbing a podniknúť tak kroky k jej zastaveniu alebo zabráneniu.

POĎAKOVANIE

Tento príspevok vznikol s podporou KEGA grantu MŠVVaŠ SR č. 065TUKE-4/2011 (50%). Tento príspevok vznikol tiež s podporou a realizáciou projektu „Rozvoj Centra informačných a komunikačných technológií pre znalostné systémy (ITMS kód: 26220120030)“ (50%).

LITERATÚRA

[1] Marketing na sociálnych sieťach [on-line] [cit 2012.4.25]. Dostupné na internete: <http://marketing.krea.sk/clanky/socialne-siete/>
[2] Mapa sveta podľa sociálnych sietí [on-line] [cit 2012.4.10]. Dostupné na internete: <http://blog.pelikan.sk/pelikan_mix/mapa-sveta-podla-socialnych-sieti/>
[3] Ako je to so sociálnymi sieťami na Slovensku? [on-line] [cit 2012.4.5]. Dostupné na internete: <http://www.uspesnaskola.sk/>
[4] On-line generácia [on-line] [cit 2011.5.5]. Dostupné na internete: <http://ucm-sk.academia.edu/NorbertVrabec/Papers/1126798/ON-LINE_GENERACIA_informacie_komunikacia_a_digitalna_participacia_mlade_e_v_informacnej_spolocnosti>
[5] MACHOVÁ, K., KRAJČ, M.: Opinion Classification in Threaded Discussions on the Web. 10th Annual International Conference Znalosti 2011, Stará lesná, 31.1.–2.2.2011, Publisher: FEI Technická univerzita Ostrava, Czech Republic, 2011, 136-147, ISBN 978-80-248-2369-0.
[6] New CISPA cybersecurity bill will censor the Web [on-line] [cit 2012.5.9]. Dostupné na internete: <http://rt.com/usa/news/cispa-bill-sopa-internet-175/>
[7] Vývoj a zmeny v používaní internetu [on-line] [cit 2012.5.2]. Dostupné na internete: <http://www.mobil.sk/clanok11612-Ste_NGUjovia_alebo_internetovi_konzervativci.htm>
[8] Komunikácia v prostredí internetu [on-line] [cit 2012.4.17]. Dostupné na internete: <www.osvetake.sk/data/Projekt-Kyber.../online_komunikacia.ppt>
[9] Problémy s dátami a metódy získavania znalostí z dát [on-line] [cit 2012.4.10]. Dostupné na internete: <http://www2.fiit.stuba.sk/~kapustik/ZS/Clanky0304/sas/index.htm>
[10] PARALIČ, J., a kol.: Dolovanie znalostí z textov. Equilibria, s.r.o., Košice, 2010, 184 s.

Návrh a simulácia riadenia pre výukový model Magnetická levitácia

Peter ŠUSTER, Martin BENETIN, Anna JADLOVSKÁ

Katedra kybernetiky a umelej inteligencie, Fakulta elektrotechniky a informatiky,
Technická univerzita v Košiciach, Slovenská republika

anna.jadlovka@tuke.sk, martin.benetin@student.tuke.sk, peter.suster@tuke.sk

***Abstrakt* — Cieľom tohto článku je navrhnúť algoritmus riadenia pre nelineárny systém Magnetická levitácia s využitím metód lineárnej syntézy - LQ riadenie s integrátorom a nelineárnej syntézy - metóda exaktnej linearizácie vstup - výstup. V článku je uvedený matematický model Magnetickej levitácie, na základe ktorého bol vytvorený nelineárny simulačný model Magnetická levitácia v prostredí Simulink. Navrhnuté algoritmy riadenia spolu so simulačným modelom Magnetickej levitácie sú implementované do jednotlivých riadiacich štruktúr s cieľom sledovať zadané referenčné trajektórie, čo je simulačne overené v prostredí Matlab/Simulink.**

***Kľúčové slová* — exaktná linearizácia, magnetická levitácia, stavové riadenie s integrátorom**

I. ÚVOD

Systém Magnetická levitácia je príkladom nelineárneho, v otvorenej slučke nestabilného systému s rýchlou dynamikou. Pre tieto vlastnosti systému Magnetickej levitácie je samotné modelovanie a hlavne návrh algoritmu riadenia veľmi obtiažne. Systém Magnetická levitácia ma však široké využitie v rôznych oblastiach priemyslu ako sú dopravné systémy, magnetické vlaky, sťahovanie kovových predmetov v oceliarskom priemysle pomocou takzvaných magnetických vločiek, trení ložísk a pod. a preto je tejto oblasti výskumu v posledných rokoch venované značné úsilie.

Pri návrhu algoritmu riadenia pre systém Magnetickej levitácie sa použilo niekoľko prístupov ako adaptívne riadenie [2] alebo robustné riadenie [3].

V tomto článku bude prezentované využitie metódy LQ riadenia s integrátorom a metódy exaktnej linearizácie vstup – výstup pri návrhu algoritmu riadenia pre simulačný model systému Magnetická levitácia, ktorého fyzikálny základ tvorí reálny model Magnetická Levitácia CE152 od firmy Humusoft [1], ktorý sa nachádza laboratóriu Kybernetiky na Katedre kybernetiky a umelej inteligencie na FEI TU Košice.

Simulačný nelineárny model Magnetickej levitácie spolu s navrhnutými algoritmami riadenia sú implementované do riadiacich štruktúr s cieľom riadenia na viac ustálených stavov obsiahnutých v definovanej referenčnej trajektórii resp. sledovanie referenčnej trajektórie, ktorá predstavuje funkciu sínus. Navrhnuté algoritmy lineárnej a nelineárnej syntézy sú simulačne overené vo vhodných riadiacich štruktúrach v programovacom jazyku Matlab/Simulink.

II. LINEÁRNE METÓDY SYNTÉZY - STAVOVÉ RIADENIE

Vo všeobecnosti je možne opísať dynamiku nelineárneho systému pomocou nelineárnych diferenciálnych rovníc v tvare[4]

$$\begin{aligned} \dot{x}(t) &= f(x,t) + g(x,t)u(t) \\ y(t) &= h(x,t) \end{aligned} \tag{1}$$

kde $x(t)$ je n - rozmerný vektor stavov, $u(t)$ je m - rozmerný vektor vstupov, $y(t)$ je m - rozmerný vektor výstupov, $f(x,t)$, $g(x,t)$ a $h(x,t)$ sú hladké nelineárne funkcie. Stavový opis, ktorý je získaný lineárnou aproximáciou nelineárnych diferenciálnych rovníc (1) v okolí zvoleného pracovného bodu x_{pb}, má nasledujúci tvar

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t) \\ y(t) &= Cx(t) + Du(t) \end{aligned} \qquad x(0) = x_{pb} \tag{2}$$

kde A je matica dynamiky, B je vstupná matica, C je výstupná matica, D je matica priamej väzby medzi vstupom a výstupom dynamického systému.

Základnou podmienkou pri návrhu stavového riadenia je aby dynamický systém (2) bol riaditeľný t.j. stav $x(t_1)$ bude dosiahnuteľný zo stavu $x(t_0)$ pomocou vstupného riadiaceho signálu $u(t)$ za konečný čas t_1. Dynamický systém n - tého rádu je riaditeľný ak matica dosiahnuteľnost má hodnosť *rank* $Q_d = n$, kde

$$Q_M = \begin{bmatrix} B & AB & A^2B & \dots & A^{n-1}B \end{bmatrix} \tag{3}$$

Po splnení podmienky riaditeľnosti dynamického systému je možné navrhnúť ako LQ riadenie s integrátorom s využitím približnej linearizácie nelineárneho systému vo zvolenom pracovnom bode x_{pb}, tak aj riadenie navrhnuté s využitím metódy exaktnej linearizácie. [7]

III. LQ RIADENIE S INTEGRÁTOROM - LQI RIADENIE

Výsledkom samotného LQ návrhu riadenia je proporcionálna stavová spätná väzba. Dôsledkom proporcionálnej spätnej väzby vzniká v uzavretom obvode pri zmenách žiadaných veličín a porúch trvalá regulačná odchýlka. Odstránenie trvalej regulačnej odchýlky sa dá zabezpečiť tak, že algoritmus riadenia bude mať integračnú zložku. Jeden zo spôsobov návrhu algoritmu riadenia s integračnou zložku je klasický LQ návrh na základe funkcionálu, v ktorom namiesto vektora $u(t)$ sa bude uvažovať jeho deriváciu $\dot{u}(t)$. Druhý spôsob návrhu algoritmu riadenia s integračnou zložku je klasický LQ návrh pre prípad rozšíreného vektora stavových veličín a toľko nových stavových veličín, pre ktoré sa vyžaduje mať integračnú zložku. [6]

Základnou podmienkou podobne ako pri LQ riadení je, aby dynamický systém (2) spĺňal podmienku riaditeľnosti (3). Výsledkom tejto metódy je optimálne stavové spätnoväzobné riadenie, ktoré zabezpečí požadované sledovanie referenčnej trajektórie (Obr. 1).

V prípade návrhu stavového riadenia s integračnou zložkou sa vychádza z stavového opisu (2) dynamického systému, pričom je potrebné navrhnúť riadenie v tvare

$$u(t) = -Kx_{LQI}(t) = -K\begin{bmatrix} x(t) & x_i(t) \end{bmatrix}^T \tag{4}$$

kde $x_i(t) = \int_0^\infty w(t) - y(t)dt$ je výstup z integrátora (Obr.1). Pri tejto metóde sa uvažuje funkcionál v tvare

$$J = \int_0^\infty ([y(t) - w(t)]^T Q_y [y(t) - w(t)] + m(t)^T Rm(t))dt \tag{5}$$

kde $w(t)$ je referenčná trajektória, Q_y, R sú symetrické kladne definitné matice a $m(t)$ je derivácia vstupu $u(t)$, t.j.

$$m(t) = \frac{du(t)}{dt} \tag{6}$$

Ďalej je potrebné zadefinovať nový stavový vektor $\bar{x}(t) = \begin{bmatrix} \dot{x}(t)^T, e(t) \end{bmatrix}^T$, kde $e(t)$ je regulačná odchýlka (Obr.1), rozšírený stavový opis dynamického systému (2) v tvare

$$\begin{bmatrix} \ddot{x}(t) \\ \dot{e}(t) \end{bmatrix} = \begin{bmatrix} A & 0 \\ C & 0 \end{bmatrix} \begin{bmatrix} \dot{x}(t) \\ e(t) \end{bmatrix} + \begin{bmatrix} B \\ 0 \end{bmatrix} m(t) \quad \Rightarrow \quad \dot{\bar{x}}(t) = \hat{A}x(t) + \hat{B}m(t) \tag{7}$$

a odpovedajúci rozšírený funkcionál v tvare

$$J = \int_0^\infty (\bar{x}(t)^T \begin{bmatrix} 0 & 0 \\ 0 & Q_y \end{bmatrix} \bar{x}(t) + m^T Rm)dt \tag{8}$$

Teraz je možné postupovať pri návrhu stavového riadenia s integračnou zložkou pre daný dynamický systém (2) ako pri návrhu klasického LQ riadenia pre ustálené riešenia Ricattiho rovnice [6]. Po splnení uvedeného postupu na stavové riadenie má $s(t)$ nasledujúci tvar

$$m(t) = -R^{-1}\hat{B}^T \hat{P}_p \bar{x}(t) = -K_x \dot{x}(t) - K_i e(t) \tag{9}$$

kde K_x, K_i a $\hat{P}_p$ sú dané takto:

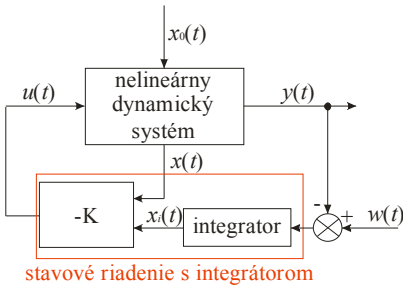
$$\begin{bmatrix} K_x & K_i \end{bmatrix} = \begin{bmatrix} R^{-1}\hat{B}^T\hat{P}_{11} & R^{-1}\hat{B}^T\hat{P}_{12} \end{bmatrix}, \quad \hat{P}_p = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix} > 0 \tag{10}$$

kde K_x je vektor zosilnení jednotlivých stavov, K_i je zosilnenie integračnej zložky a $\hat{P}_p$ je symetrická kladne semidefinitná matica, ktorá sa získa ako riešenie maticovej Riccattiho algebraickej rovnice v tvare

$$\hat{P}_p\hat{A} + \hat{A}^T\hat{P}_p - \hat{P}_p\hat{B}R^{-1}\hat{B}^T\hat{P}_p = \begin{bmatrix} 0 & 0 \\ 0 & Q_y \end{bmatrix} \tag{11}$$

Po integrácii rovnice (9) má výsledné stavové riadenie s integračnou zložkou tvar

$$u(t) = -K_x x(t) + K_i x_i(t) \tag{12}$$



Obr. 1 Riadiaca štruktúra stavového riadenia - LQI riadenie

V tejto časti bol uvedený všeobecný popis lineárnej metódy syntézy - LQ riadenie s integrátorom, na základe ktorého a s využitím približnej linearizácie bol nahrnutý výsledný zákon optimálneho stavového riadenie s integračnou zložkou (12) s cieľom riadiť simulačný nelineárny model Magnetická levitácia.

IV. NELINEÁRNE METÓDY SYNTÉZY - METÓDA EXAKTNEJ LINEARIZÁCIE VSTUP - VÝSTUP

V tejto časti bude uvedený všeobecný popis metódy exaktnej linearizácie v riadení SISO systémov ako je napríklad Magnetická levitácia. Táto metóda spočíva vo výpočte Lie derivácií, určení relatívneho rádu nelineárneho systému, lokálnej stavovej transformácie súradníc a zo vstupnej transformácie, preto najskôr budú zadefinované niektoré základné pojmy. [4]

Predpokladom je nelineárny dynamický systém zapísaný v afinnom tvare (1). Pre skalárnu funkciu $\lambda(x) = \lambda(x_1, x_2, ..., x_n)$ vektorovej premennej $x(x_1, x_2, ..., x_n)$ a n - rozmernú vektorovú funkciu

$$f(x) = \begin{pmatrix} f_1(x_1, x_2, ..., x_n) \\ f_2(x_1, x_2, ..., x_n) \\ ... \\ f_n(x_1, x_2, ..., x_n) \end{pmatrix} \tag{13}$$

je možné definovať novú skalárnu funkciu premennej x nazývanú Lieho operátor, ktorá sa označuje $L_f\lambda(x)$, pričom

$$L_f\lambda(x) = L_f\lambda(x_1, x_2, ..., x_n) = \sum_{i=1}^{n} \frac{\partial\lambda}{\partial x_i} f(x_1, x_2, ..., x_n) \tag{14}$$

Po zadefinovaní, že $\frac{\partial\lambda}{\partial x} = (\frac{\partial\lambda}{\partial x_1}, \frac{\partial\lambda}{\partial x_2}, ..., \frac{\partial\lambda}{\partial x_n})$ možno funkciu $L_f\lambda(x)$ jednoducho vyjadriť ako

$$L_f\lambda(x) = \frac{\partial\lambda}{\partial x} f(x) \tag{15}$$

Nová funkcia $L_f\lambda(x)$ sa nazýva aj derivácia $\lambda(x)$ pozdĺž $f(x)$. Možné je aj opakované použitie tejto operácie. Napríklad derivovaním najskôr pozdĺž $f(x)$ a potom pozdĺž $g(x)$ je možné skonštruovať funkciu v tvare

$$L_g L_f \lambda(x) = \frac{\partial L_f}{\partial x} g(x) \tag{16}$$

Derivovaním $\lambda(x)$ k – krát pozdĺž $f(x)$ je možné rekurzívne definovať funkciu

$$L_f^k \lambda(x) = \frac{\partial L_f^{k-1} \lambda}{\partial x} f(x) \tag{17}$$

Pomocou Lie operátora je možné zadefinovať pojem relatívneho rádu systému. Daný nelineárny systém (1) má relatívny rád r v bode x_0, ak pre všetky x z okolia x_0 a pre všetky $k < r-1$ platí, že výraz $L_g L_f^k \lambda(x) = 0$, pričom pre $k = r-1$ platí $L_g L_f^{r-1} \lambda(x) \neq 0$.

Predpokladá sa, že nelineárny systém (1) má relatívny rád r v bode x_0, pričom $r = n$, kde n je rád nelineárneho systému. Následne rekurzívnym výpočtom Lie operátorov je možné získať lokálnu stavovú transformáciu súradníc $\Phi(x)$, ktorá je definovaná nasledovne

$$\Phi(x) = \begin{pmatrix} \Phi_1(x) \\ \Phi_2(x) \\ \dots \\ \Phi_n(x) \end{pmatrix} = \begin{pmatrix} \lambda(x) \\ L_f \lambda(x) \\ \dots \\ L_f^{r-1} \lambda(x) \end{pmatrix} \tag{18}$$

Zobrazenie $\Phi(x)$ má regulárnu Jacobiho maticu v bode x_0, a preto možno toto zobrazenie považovať za lokálnu stavovú transformáciu súradníc v okolí bodu x_0. Nech prvá nová súradnica je rovná $z_1 = y = h(x)$ potom stavový opis systému v nových súradniciach je v tvare

$$\begin{aligned} \dot{z}_1 &= \frac{\partial \Phi_1(x)}{\partial x}\frac{\partial dx}{\partial dt} = \frac{\partial h(x)}{\partial x}\frac{\partial dx}{\partial dt} = L_f h(x) = \Phi_2(x) = z_2 \\ &\vdots \\ \dot{z}_{r-1} &= L_f^{r-1} h(x) = \Phi_r(x) = z_r \\ \dot{z}_r &= L_f^r h(x) + L_g L_f^{r-1} h(x) u \end{aligned} \tag{19}$$

Na získanie lineárnej formy nelineárneho systému sú potrebné tri kroky

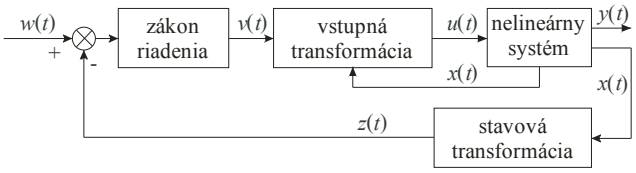
1. určenie relatívneho rádu r systému, pre pokračovanie v syntéze musí platiť $r = n$, kde n je rád nelineárneho systému
2. lokálna stavová transformácia súradníc podľa rovníc (18)
3. definovanie vstupnej transformácie nelineárneho systému z poslednej rovnice sústavy rovníc (19) pri substitúcii $\dot{z}_r = v$ v tvare

$$u = \frac{1}{L_g L_f^{r-1} h(x)} (v - L_f^r h(x)) \tag{20}$$

kde v predstavuje nový vstup do systému opísaného sústavou rovníc (19). Nový transformovaný systém sa správa ako séria za sebou zapojených integrátorov. Tento transformovaný systém je v stavovom opise vyjadrený v tvare

$$\begin{aligned} \dot{z} &= \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix} z + \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} v \\ y &= \begin{pmatrix} 1 & 0 & \dots & 0 \end{pmatrix} z \end{aligned} \tag{21}$$

Na lineárny systém (21) je možné aplikovať lineárne metódy syntézy, ako napríklad stavové optimálne riadenie s využitím princípu LQ alebo LQI popísané v časti II tohto článku a podobne, pomocou ktorých sa navrhne zákon riadenia, ktorý zabezpečí požadované správanie sa nelineárneho systému (1). Získané transformácie z metódy exaktnej linearizácie ako aj navrhnuté lineárne riadenie je následne implementované do riadiacej štruktúry (Obr. 2).

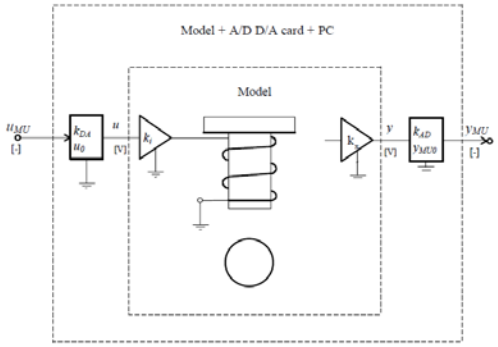


Obr. 2 Bloková schéma implementácie spätnoväzobnej exaktnej linearizácie s LQI riadením do riadiacej štruktúry

Na základe uvedeného popisu metódy exaktnej linearizácie bol navrhnutý programový modul s využitím Symbolického toolboxu v prostredí Matlab/Simulink, ktorého výsledkom boli výsledné transformácie a tiež aj zákon riadenia, pre ktorého návrh bola použitá metóda LQI riadenia. [8]

V. VERIFIKÁCIA ALGORITMOV RIADENIA NA SIMULAČNOM MODELY MAGNETICKÁ LEVITÁCIA

Model magnetickej levitácie CE152 od spoločnosti Humusoft je znázornený na Obr. 3 a demonštruje problémy riadenia späté s nelineárnym nestabilným systémom. Jedná sa o SISO systém, kde vstupom je napätie prúdového zosilňovača *u* a výstupom je poloha guľôčky *r* v magnetickom poli. Systém pozostáva z cievky, ktorá levituje oceľovú guľôčku v magnetickom poli. Pozícia guľôčky *y* sa získava pomocou indukčného lineárneho pozičného senzora spojeného s A/D prevodníkom. Cievka je budená prúdovým zosilňovačom, ktorý je spojený s D/A prevodníkom.[1]



Obr. 3 Bloková schéma modelu magnetickej levitácie

Stavový opis (afinny tvar) modelu Magnetická levitácia, ktorý je potrebný pri použití metódy exaktnej linearizácie ako aj pri LQI riadení má nasledujúci tvar

$$\begin{bmatrix}\dot{x}_1\\ \dot{x}_2\\ \dot{x}_3\end{bmatrix} = \begin{bmatrix} x_2 \\ \dfrac{x_3^2 k_c}{m_k (x_0 - x_1)^2} - g - \dfrac{k_{fv}}{m_k} x_2 \\ -\dfrac{x_3}{T_a} \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \dfrac{k_i k_{da}}{T_a} \end{bmatrix} u_{MU} \qquad (22)$$

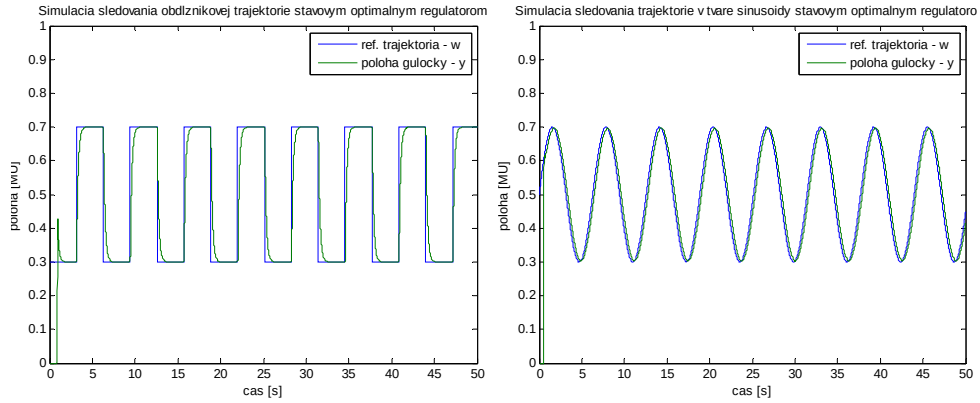
$$y_{MU} = \begin{bmatrix} k_x k_{ad} x_1 \end{bmatrix}$$

kde fyzikálne veličiny a parametre sú uvedené v Tabuľke 1.

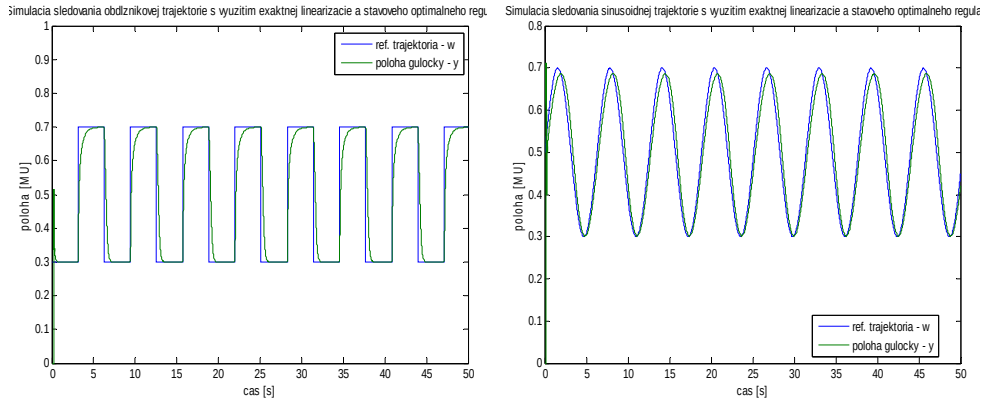
Tabuľka. 1
Fyzikálne veličiny a parametre modelu Magnetická levitácia

Fyzikálne veličiny:	m_k - hmotnosť guľôčky [kg],
$x = [x_1, x_2, x_3] = [r, \dot{r}, i]$	x_0 - ofset polohy guľôčky [m],
r - poloha guľôčky [m],	g - gravitačné zrýchlenie [ms^{-2}],
$\dot{r}$ - rýchlosť guľôčky [ms^{-1}],	k_{fv} - koeficient tlmenia [Nm^{-1}s^{-1}],
i - prúd pretekajúci cievkou [A],	k_i - zosilnenie cievky a zosilňovača [A/V],
Parametre:	T_a - časová konštanta cievky a zosilňovača [s],
u_{MU} - vstup D/A prevodníka [MU],	k_x - zosilnenie pozičného senzora [Vm^{-1}],
y_{MU} - výstup A/D prevodníka [V],	k_{da} - zosilnenie D/A prevodníka [V/MU],
k_c - konštanta cievky,	k_{ad} - zosilnenie A/D prevodníka [MU/V]

Na základe stavového opisu (22) bol naprogramovaný nelineárny simulačný model Magnetickej levitácie v jazyku Matlab/Simulink, ktorý spolu s navrhnutými algoritmami riadenia, bol implementovaný do jednotlivých riadiacich štruktúr (Obr. 1)(Obr. 2), s cieľom zabezpečiť požadovanú polohu guľôčky v magnetickom poli cievky. Výsledné grafy sledovania referenčných trajektórii pri použití lineárnej a nelineárnej syntézy sú na Obr.4 a Obr.5.



Obr. 4 Výsledné priebehy sledovania referenčnej trajektórie výstupom simulačného modelu Magnetická levitácia - LQI riadenie



Obr. 5 Výsledné priebehy sledovania referenčnej trajektórie výstupom simulačného modelu Magnetická levitácia - exaktná linearizácia + LQI riadenie

VI. ZÁVER

V tomto článku bol prezentovaný návrh algoritmov riadenia pre nelineárny simulačný model Magnetická levitácia s využitím lineárnej metódy syntézy - LQ riadenie s integrátorom a nelineárnej metódy syntézy - metóda exaktnej linearizácia vstup – výstup. Navrhnuté zákony riadenia spolu so simulačným nelineárnym modelom bolo implementované do riadiacich štruktúr a simulačné overené v prostredí Matlab/Simulink. Výsledné priebehy ukazujú že výstup z modelu sleduje referenčné trajektórie definované buď ako skokové zmeny medzi dvoma ustálenými stavmi alebo trajektóriou, ktorá predstavuje funkciu sínus, a preto je možné považovať dané prístupy za vhodné na riešenie problému riadenia pre nelineárny model Magnetická levitácia. Získane poznatky a simulačné výsledky z oblasti metód lineárnych a nelineárnych syntéz budú použité pre návrh riadenia reálneho modelu Magnetická levitácia a tiež pre aplikačné využitie na cvičeniach v predmetoch Optimálne a nelineárne systémy, a Riadenie a umelá inteligencia.

POĎAKOVANIE

Tento článok bol vytvorený realizáciou vedeckého projektu Vega č. 1/0286/11 Grantovej agentúry SR pod názvom Dynamické hybridné architektúry v multiagentových sieťových riadiacich systémoch (50%) a projektu Rozvoj Centra informačných a komunikačných technológií pre znalostné systémy (ITMS kód: 26220120030) na základe podpory operačného programu Výskum a vývoj financovaného z Európskeho fondu regionálneho rozvoja (50%).

LITERATÚRA

[1] Humusoft Praha: CE 152 Education Manual Magnetic levitation
[2] SHAFIQ, M, AKHTAR, S.: Inverse Model Based Adaptive Control of Magnetic Levitation System, 5th Asian Control Conference. 2004.
[3] JACKO, Ľuboš, HYPIUSOVÁ, Mária: Návrh robustného riadenia magnetickej levitácie na základe vety o stabilite na hrane kvádra neurčitosti, ŠVOČ 2007 sekcia: Automatizované systémy riadenia, STU Bratislava, 2007.
[4] HUBA, Mikuláš.: Nelineárne systémy, Bratislava: STU, Lokálne stredisko dištančného vzdelávania 2003. ISBN 80-227-1908-0
[5] SASTRY, Shankar: Nonlinear Systems - Analysis, Stability and Control. Springer-Verlag New York Inc., 1999. ISBN 0-387-98513-1
[6] FIKAR, Miroslav: Dynamická optimalizácia procesov, STU Bratislava, 2010, ISBN 978-80-89316-08-3.
[7] KROKAVEC, Dušan, FILASOVÁ, Anna: Diskrétne systémy, Elfa, Košice, 2008.
[8] ŠUSTER, Peter, JADLOVSKÁ, Anna: Modeling and control design of magnetic levitation system, In: SAMI 2012 : 10th IEEE Jubilee International Symposium on Applied Machine Intelligence and Informatics : proceedings : Herľany, Slovakia, January 26-28, 2012. - Budapest : IEEE, 2012. - ISBN 978-1-4577-0195-5. - S. 295-299
[9] BENETIN, Martin: Výukové programové moduly pre analýzu a syntézu modelov fyzikálnych systémov, Diplomová práca, Košice: Technická univerzita v Košiciach, Fakulta elektrotechniky a informatiky, 2009.

Možnosti nasadenia SaaS v konkrétnej firme

Daniela TOMKOVÁ, Adela TUŠANOVÁ, Ján PARALIČ

Katedra kybernetiky a umelej inteligencie, Fakulta elektrotechniky a informatiky,
Technická univerzita v Košiciach, Slovenská republika

daniela.tomkova@student.tuke.sk, adela.tusanova@tuke.sk, jan.paralic@tuke.sk

***Abstrakt* – V článku sa venujeme problematike SaaS (Softvér ako služba - Software as a Service) a možnostiam nasadenia tohto typu aplikácie v konkrétnej stavebnej firme. Uvádzame definíciu základných pojmov, charakteristiku vlastností, jej výhody a obmedzenia. Informujeme o existujúcich aplikáciách nasadených formou SaaS doma i vo svete. Uvádzame krátky popis analýzy konkrétnej firmy vo vzťahu k možnostiam nasadenia konceptu SaaS. Navrhujeme model praktického nasadenia prípravy a sledovania stavebnej zákazky s podrobným popisom technického riešenia.**

***Kľúčové slová* - Cloud Computing, softvér ako služba, Enterprise Resource Planning, časový plán stavebnej zákazky**

I. ÚVOD

Nachádzame sa v období veľkých a dramatických technologických a ekonomických zmien, v rámci ktorých hrajú informačné systémy a technológie stále väčšiu rolu a vo veľkej miere ovplyvňujú významné oblasti ekonomiky. Čoraz viac odvetví v stále väčšej miere využíva softvér dodávaný vo forme on-line služby v globálnom meradle.

Dnes viac ako dve miliardy ľudí používa širokopásmové pripojenie k internetu, pričom pred desiatimi rokmi to bolo len 50 miliónov. V najbližších desiatich rokoch sa očakáva, že najmenej 5 miliárd ľudí na celom svete bude vlastniť smartfóny s okamžitým a nepretržitým prístupom na internet.

Spoločnosti v každom odvetví zaregistrovali nastupujúcu softvérovú revolúciu, vyznačujúcu sa predovšetkým poskytovaním softvérových aplikácií on-line cez internet [1].

Prístup k aplikáciám, ktoré sú umiestnené na internete, teda inde ako na lokálnom počítači, zabezpečuje stále viac používaná koncepcia nazývaná cloud computing [2].

Základná myšlienka cloud computingu spočíva v tom, že infraštruktúra aplikácie pracuje priamo na webe - od jednoduchého softvéru až po kompletné operačné systémy. Zákazník teda nepotrebuje drahý a výkonný hardware na náročné výpočtové operácie. Nemusí sa starať o aktualizáciu systému, dokonca už nemusí kupovať žiadny softvér, jednoducho si ho prenajme od poskytovateľa cloudových služieb [3]. NIST (National Institute of Standards and Technology) definuje tri modely cloudových služieb: SaaS - Software as a Servis (softvér ako služba), PaaS - Platform as a Service (platform ako služba) a IaaS - Infrastructure as a Service (infraštruktúra ako služba) [24]. Predmetom nášho záujmu je SaaS.

II. SOFTVÉR AKO SLUŽBA

A. Koncept – Softvér ako služba

Softvér ako služba, angl. Software as a Service (SaaS), je poskytovanie softvéru formou prenájmu softvérovej aplikácie prístupnej prostredníctvom zabezpečeného miesta na internete. Uskutočňuje sa v rámci koncepcie, ktorú nazývame cloud computing. Zákazník nemusí aplikáciu ani inštalovať ani aktualizovať, ani udržiavať. Aplikácia je celosvetovo prístupná a okamžite použiteľná.

Termín služby v cloud computingu označuje možnosť práce s opakovane použiteľnými a jemne štruktúrovanými komponentmi v rámci siete dodávateľa [2].

Koncepcia SaaS je veľmi atraktívna, má však jednu nevýhodu: aj aplikácia aj dáta sú pod úplnou kontrolou dodávateľa softvéru. Bezpečnosť dát a nepretržitý prístup k aplikácii patria

medzi hlavné priority nasadenia SaaS. Vzhľadom na citlivosť tejto otázky si dodávatelia všetkých vrstiev cloudu dávajú na bezpečnosti mimoriadne záležať, aby si získali bezpodmienečnú dôveru svojich klientov.

B. Výhody a obmedzenia SaaS

Nasadzovaniu SaaS v segmente podnikových aplikácií sa darí vo svete už niekoľko rokov a tento trend si začínajú všímať aj manažéri našich podnikov. SaaS aplikácie totiž riešia veľa problémov, ktoré dlhodobo trápia tradičné podnikové systémy, ako sú veľké licenčné poplatky platené vopred, dlhodobá implementácia, komplikovaná aktualizácia a slabá IT podpora [5].

Výhody riešenia "softvér ako služba" oproti "on-premise software", t.j. oproti tradičnému licenčnému softvéru možno stručne zhrnúť takto: tradičný softvér vyžaduje inštaláciu z CD, resp. DVD, dlhú implementáciu do podnikových procesov (často celé mesiace), vysoké licenčné poplatky na začiatku, nová funkcionalita si vyžaduje nový upgrade, je nevyhnutná externá, resp. interná IT podpora a údržba, softvér má špecifikované nároky na hardware, bezpečnosť a ochrana dát je závislá na internom oddelení IT alebo na samotnom užívateľovi.

Pri riešení typu SaaS nie je potrebná žiadna inštalácia, implementácia je rýchla a flexibilná (dni, týždne), prístup k aplikácii je jednoduchý a rýchly odkiaľkoľvek prostredníctvom webového prehliadača, poplatky sú relatívne nízke - platby iba za skutočnú spotrebu, nová funkcionalita je dodávaná automaticky, typická je škálovateľnosť (elasticita), t.j. schopnosť dynamicky sa prispôsobovať meniacej sa záťaži, taktiež sú výrazne znížené náklady na IT servis a v neposlednom rade bezpečnosť je zaručená technológiou a infraštruktúrou svetových dodávateľov.

Riešenia typu SaaS vyhľadávajú hlavne malé a stredné firmy. Väčšie podniky sú opatrnejšie. Obvykle totiž pracujú v prostredí prísnejších interných, resp. externých predpisov a preto kladú väčší dôraz na spracovanie dôverných dát a na tzv. redundanciu, t.j. na spustenie náhradného riešenia, ak hlavné riešenie zlyhá. Navyše s používaním SaaS je spojené riziko, že predplatené softvérové služby zdražejú väčším dopytom po nich. Dôležitou témou na riešenie je potreba prispôsobiť a integrovať dodanie SaaS s existujúcimi programovými systémami. Medzi obmedzujúce faktory koncepcie SaaS zahŕňame taktiež stratu výhradnej kontroly nad podnikovými dátami a procesmi [6], potenciálne výpadky pripojenia k internetu a nedoriešené právne otázky v tejto oblasti.

Z vyššie uvedených dôvodov je teda zrejmé, že výber dodávateľa aplikácie SaaS je pre firmu mimoriadne citlivý. Dodávatelia SaaS musia riešiť okrem funkcionality dodávaného riešenia aj zabezpečenie ďalších nevyhnutných technických požiadaviek.

C. Minimálne technické požiadavky na SaaS

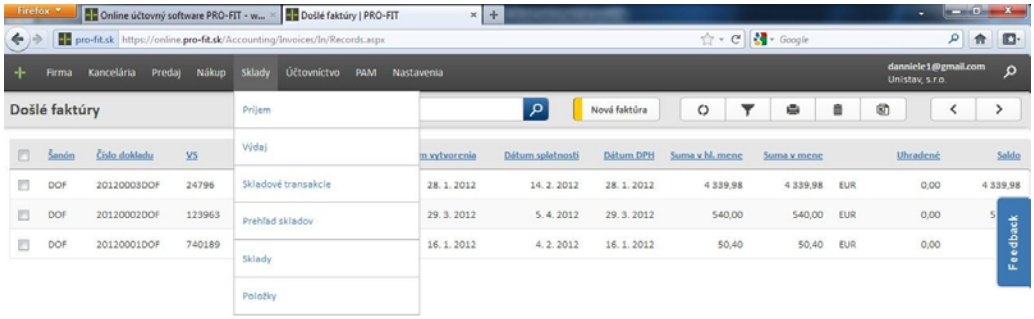
Aplikácia SaaS je technologicky navrhnutá tak, aby ju mohlo využívať množstvo zákazníkov na jednej zákaznícke inštancii (Multi_Tenant architektúra) s bezpečným mechanizmom dátového prístupu [4]. Nevyhnutným predpokladom dodávky modelu je samoobslužný prihlasovací systém s automatickým prístupom k zdrojom infraštruktúry ako aj podpora transparentných aktualizácií a opráv. Pre kvalitu prevádzky aplikácie má význam monitorovanie výkonu SaaS. Súčasťou aplikácie je taktiež systém pre meranie jej využitia, ktorý je prepojený so systémom fakturácie spotrebovaných služieb.

III. ANALÝZA NASADENIA SAAS V KONKRÉTNEJ FIRME

Možnosti nasadenia SaaS sme skúmali v stavebnej firme Unistav, s.r.o. Prešov s približne 150 zamestnancami [23]. Oboznámili sme sa s činnosťou jednotlivých oddelení a s organizačnou štruktúrou. Firma používa líniovú organizačnú štruktúru, ktorá je pri realizácii stavebných zákaziek kombinovaná s tzv. projektovou koordináciou. Vo firme sú v prevádzke dva informačné systémy: program na vedenie finančného účtovníctva, miezd, personalistiky a materiálno technického zabezpečenia IFOsoft [7] a program na tvorbu stavebno-montážnych rozpočtov, kalkulácií, harmonogramov a na prípravu stavby Systematic Kalkulus [8]. Tieto systémy automatizujú iba jednotlivé obchodné funkcie, nie však proces podnikania ako celku.

Manažment si uvedomuje, že firma potrebuje informačný systém pre správu a koordináciu všetkých firemných zdrojov, pracovísk a podnikových funkcií prostredníctvom zdieľaných dátových úložísk [9], t.j. systém typu ERP (Enterprise resource planning). V minulosti už firma komunikovala s viacerými dodávateľmi tradičných licenčných ERP informačných systémov. My sme uskutočnili prieskum ERP systémov, dodávaných formou SaaS. Trh ERP SaaS

pozostáva z riešení veľkých hráčov ako sú SAP [12], Oracle [13], Microsoft Dynamics Navision [14], ktorým úspešne začínajú konkurovať menšie systémy ako napríklad NetSuite [15], Epicor [16], Infor [17], Plex Systems [18] a ďalšie. Uvedené ERP systémy v prevedení SaaS nie sú lokalizované v slovenskom, resp. českom jazyku, čo ich v súčasnosti a priori diskvalifikuje na nasadenie v slovenskej firme. Existujúce lokálne systémy dodávané ako SaaS sa orientujú zatiaľ iba na účtovné systémy. Možnosť nasadenia aplikácie SaaS sme otestovali na konkrétnom účtovnom systéme slovenskej výroby PRO-FIT [19]. Test prebiehal na ekonomickom oddelení firmy s aktuálnymi účtovnými dokladmi.



Obr. 1 Náhľad na účtovný doklad testovanej aplikácie SaaS.

Výsledky testu možno zhrnúť do nasledovných bodov:

- rýchla samoobslužná registrácia firmy v systéme,
- bez akejkoľvek inštalácie programu sa v priebehu pár minút začalo účtovať (bola použitá preddefinovaná účtovná osnova a nezadávali sa počiatočné stavy účtov),
- nulová počiatočná investícia je irelevantná, vzhľadom k tomu, že firma už má zakúpené licencie na používaný účtovný systém,
- cena za prevádzku testovanej aplikácie SaaS je vzhľadom na veľké počty dokladov vyššia ako cena upgrade existujúceho softvéru,
- z hľadiska definície ERP testovaná aplikácia SaaS pokrýva iba finančné riadenie, časť riadenia ľudských zdrojov a malú časť systému riadenia predaja,
- boli vznesené nasledovné otázky zo strany používateľov na dodávateľa testovanej SaaS aplikácie:
  - reakcia na špecifické podmienky firmy
  - kvalifikovaná pomoc pri integrácii s ostatnými procesmi firmy
  - zabezpečenie dátovej prepojiteľnosti s iným softvérom
  - migrácia dát existujúceho informačného systému do SaaS aplikácie
  - prístupnosť dát v prípade zmeny dodávateľa aplikácie SaaS
  - archivácia dát
  - bezpečnostné hľadiská

Vzhľadom na vyššie uvedené skutočnosti došlo k rozhodnutiu nenasadiť túto aplikáciu SaaS do prevádzky.

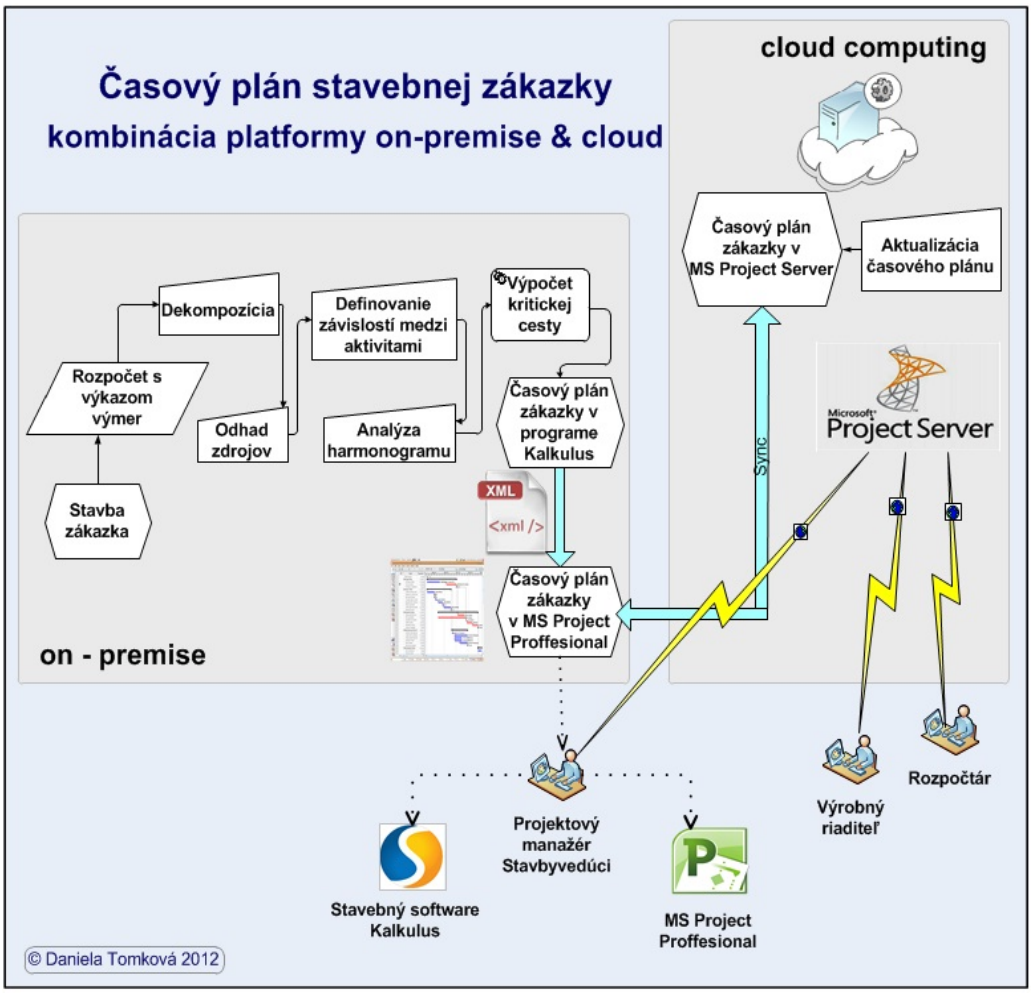
Návrh modelu prípravy a sledovania stavebnej zákazky

Hľadaním možností nasadenia SaaS v konkrétnej firme sme dospeli k poznaniu, že nazerať na firmu „zvonku“ prostredníctvom už existujúcich technických riešení nevedie k uspokojivým výsledkom. Rozhodli sme sa preto preskúmať firmu „zvnútra“, t.j. zmapovať takú oblasť, kde prebiehajú procesy neoptimálne bez informačnej podpory softvérových nástrojov. Takouto oblasťou je „Riadenie výroby“. Kým proces prípravy stavby je vo firme plne automatizovaný, zákaznícky orientovaný a prináša firme zisk, proces sledovania a riadenia zákazky prebieha dispečerským spôsobom prakticky bez podpory automatizačných prostriedkov.

Vzhľadom k tomu, že každá stavba je geograficky vzdialená od sídla firmy, prístup riadiaceho personálu stavby k spoločným firemným dátam je zdĺhavý a ich vzájomné zdieľanie s firemným manažmentom a s ostanými úsekmi v reálnom čase je prostredníctvom tradičného softvéru prakticky nemožné. Táto situácia je predurčená na použitie cloudového riešenia.

Navrhli sme model, ktorý efektívnym spôsobom dokáže kombinovať funkčnosť existujúceho tradičného softvéru (on - premise) pre prípravy zákazky s aplikáciou bežiacou v cloude formou SaaS pre sledovanie a riadenie zákazky – časť časový plán.

Schéma navrhovaného modelu je znázornená vo forme diagramu procesov na obr. 2.



Obr. 2 Model sledovania a riadenia stavebnej zákazky – časť časový plán ako kombinácia konceptu on-premise & SaaS.

A. Proces prípravy zákazky

Proces prípravy zákazky začína už pri tvorbe ponukového rozpočtu. Stavebná firma oceňuje projektantom predložený výkaz výmer. Ak výkaz výmer nie je k dispozícii, rozpočet vytvára na základe zamerania na stavbe. Výkaz výmer je zoznam cenníkových položiek s vyčíslenými množstvami (výmermi), ktoré popisujú technické a technologické riešenie stavby.

Technologický popis položky je reprezentovaný kalkulačným rozborom, t.j. onormovanou spotrebou materiálov, práce jednotlivých profesií a strojov, potrebných na zhotovenie jednej mernej jednotky príslušnej položky.

Tieto údaje slúžia:

- na výpočet ponukovej kalkulovanej ceny
- na výpočet limitov materiálu, profesií a strojov
- ako návrh zdrojov v časovom pláne.

Cenníkové (technologické) položky slúžia nielen ako položky rozpočtu, ale pri tvorbe časového plánu sa stávajú činnosťami (tasks).

Proces prípravy zákazky v predmetnej stavebnej firme je automatizovaný programovým systémom Kalkulus [8], ktorý je tradičnou softvérovou aplikáciou. Jeho výsledkom je výkaz výmer, ponukový rozpočet, podklady pre prípravu výroby (príslušné limity) a výrobný harmonogram stavebnej zákazky.

B. Proces tvorby a aktualizácie časového plánu

Ak firma získa zákazku a podpíše sa zmluva, stavbu preberie stavbyvedúci, ktorý má na starosti riadenie stavby v súčinnosti s ďalšími útvarmi firmy. Naštuduje dokumentáciu, vypracuje technologický postup a rozvrhne zabezpečenie potrebných materiálov, nasadenie remeselníkov, strojov a subdodávateľov v správnej postupnosti a na správne miesto. Vytvorený

harmonogram prerokuje so všetkými zainteresovanými stranami. Plán je väčšinou napätý a tlak na dodržanie termínu je veľký [20]. Ak v priebehu výstavby nastane sklz niektorej činnosti alebo dodávky v dôsledku počasia, nedisciplinovanosťou subdodávateľa alebo vlastnou chybou riadenia, vzniká problém. Stavbyvedúci po vyriešení situácie musí uskutočniť zmenu časového plánu a upovedomiť o danej zmene zainteresované strany. Je to významný a často sa opakujúci moment v riadení stavby. Navyše je špecifický tým, že stavba je geograficky vzdialená od sídla firmy. V súčasnosti sa takáto zmena nevykonáva, čo sa prejaví v neefektívnosti riadenia.

C. Automatizácia procesu tvorby a aktualizácie časového plánu

Pre proces aktualizácie časového plánu a pružné riadenie stavby sme navrhli riešenie založené na kombinácii tradičného softvéru (on-premise) a softvéru prevádzkovaného ako služba (SaaS). Ako tradičné (on-premise) programy v tomto procese slúžia programový systém Kalkulus a Microsoft Project Professional 2010. Ako SaaS v tomto procese slúži Microsoft Project Server 2010.

Microsoft Project Server 2010 a Microsoft Project Professional 2010 vytvárajú spolu platformu Microsoftu pod názvom Microsoft Enterprise Project Management (EPM) [21]. Microsoft Project Server 2010 je vyvinutý na báze SharePoint Server 2010.

Užívateľmi systému s oprávneným prístupom do cloudu k aplikácii Microsoft Project Server 2010 sú: stavbyvedúci alebo projektový manažér stavby, cenové oddelenie alebo oddelenie pre prípravu stavby, výrobný riaditeľ alebo riaditeľ firmy, prípadne ďalšie zainteresované útvary firmy.

Stavbyvedúci si prevezme podklady spracované v programe Kalkulus v rámci prípravy stavby, vytvorí si výrobný harmonogram v programe Kalkulus a vyexportuje ho do súboru typu XML. Tento súbor si importuje do MS Project Professional 2010, ktorý následne dokáže otvoriť a synchronizovať pomocou Microsoft Project Server 2010.

Keďže Microsoft Project Server 2010 beží v cloude, stavbyvedúci môže cez internetový prístup časový plán kedykoľvek aktualizovať, pričom ďalší používatelia s oprávneným prístupom ho môžu rovnakým spôsobom využívať pre ďalšie plnenie riadiacich úloh. Microsoft Project Server 2010 odporúčame prevádzkovať napríklad na Project Connect Online 2010 spoločnosti Project Hosts [22] ako SaaS za cca 50 $/užívateľa/na mesiac (2 užívatelia minimum).

V súčasnosti prebieha príprava skúšobnej prevádzky navrhovaného riešenia pre konkrétnu stavbu obchodného centra.

IV. ZÁVER

Skúmaním možností nasadenia konceptu SaaS (Softvér ako služba - Software as a Service) v konkrétnej firme sa potvrdil záujem manažmentu o túto oblasť. Došli sme k spoločnému záveru, že mapovaním interných procesov je možné nájsť vo firme také oblasti, kde riešenia typu SaaS nájdu svoje opodstatnenie. Našim prínosom bolo zmapovanie procesu prípravy, sledovania a riadenia stavebnej zákazky ako aj návrh a realizácia kombinovaného riešenia tradičného softvéru a softvéru typu SaaS. Toto riešenie je okamžite aplikovateľné v praxi. Ukázalo sa, že zavádzanie novej technológie nie je bezproblémové a nie je možné okamžite opustiť tradičné koncepty. Vzhľadom na existenciu štandardov pre výmenu dát je možné uvažovať aj o efektívnej spolupráci tradičných aplikácií s aplikáciami typu SaaS. Ambícia manažmentov využiť koncepciu SaaS môže preveriť funkčnosť procesov v rámci ich firiem. Je však zrejmé, že trend vstupu internetu do ostrých podnikových aplikácií sa už nedá zastaviť.

ACKNOWLEDGMENT

Táto práca bola podporovaná Vedeckou grantovou agentúrou Ministerstva školstva, vedy, výskumu a športu Slovenskej republiky a Slovenskej akadémie vied na základe zmluvy č. 1/1147/12 (50%) a realizáciou projektu Rozvoj Centra informačných a komunikačných technológií pre znalostné systémy (ITMS kód: 26220120030) na základe podpory operačného programu Výskum a vývoj financovaného z Európskeho fondu regionálneho rozvoja (50%).

LITERATÚRA

[1] ANDREESEN, Marc: Why Software Is Eating The World [online]. The World Wall Street Journal, 2011. [s.a.]. [cit. 2011-10-15]. Dostupné na internete: <http://online.wsj.com/article/SB10001424053111903480904576512250915629460.html>.

[2] VELTE, Anthony.T.: Cloud Computing Praktický průvodce. Computer Press, 2011. 344s. ISBN 978-80-251-3333-0.
[3] SETIKOVSKÁ, Blanka: Cloud Computing: Diplomová práca [online]. Praha: ČVUT FEL, 2010. [s.a.]. [cit. 2011-10-15]. Dostupné na internete: <http://support.dce.felk.cvut.cz/mediawiki/images/c/c8/Dp_2010_setikovska_blanka.pdf>.
[4] Aiken, Larry: Scaling Your SaaS Business, Multi-Tenancy is the Key [online]. Cloudbook Journal Vol 2 Issue 4, 2011. [s.a.]. [cit. 2011-11-08]. Dostupné na internete: <http://www.cloudbook.net/resources/stories/scaling-your-saas-business>.
[5] TAMÁŠ, Miroslav: Výhody SaaS řešení [online]. ARC Consulting, 2010. [s.a.]. [cit. 2012-01-07]. Dostupné na internete: <http://arc-consulting.sk/produkty/adaptive-planning/features/saas-advantages/ >.
[6] GFI Software: ON-PREMISE VS. CLOUD-BASED SOLUTIONS [online]. 2010. [s.a.]. [cit-2012-01-13]. Dostupné na internete: <http://www.gfi.com/whitepapers/Hybrid_Technology.pdf>.
[7] http://www.ifosoft.sk/
[8] http://www.systematic.sk/
[9] ŠVEC, Petr: Cloud computingu v praxi: cloud a bezpečnost – 2. díl [online]. 2010. [s.a.]. [cit-2012-01-14]. Dostupné na internete: <http://www.itbiz.cz/cloud-computing-v-praxi-cloud-a-bezpecnost-2>.
[10] Putting ERP in the cloud, ComputerWeekly.com, 8.apríl 2010
[11] ww.computerweekly.com/Articles/2010/04/08/240850/Putting-ERP-in-the-cloud.htm [online] 26.12.2011
[12] http://www.sap.com/sk/index.epx
[13] http://www.oracle.com/us/solutions/midsize/enterprise-class-erp-1375433.html
[14] http://www.microsoft.com/en-us/dynamics/erp-nav-overview.aspx
[15] The #1 Cloud ERP / Financials Software Suite. URL:www.netsuite.com [online] 27.11.2011
[16] http://www.epicor.com/Cloud/Pages/Cloud-Manufacturing.aspx
[17] http://www.infor.com/product_summary/erp/sl/
[18] http://www.plex.com/
[19] http://www.pro-fit.sk [online] 1.12.2011
[20] Koncept riadenia stavebnej firmy, interný materiál Systematic, 2011
[21] Project Server 2010 with SharePoint Server 2010 architecture (overview) URL: http://technet.microsoft.com/en-us/library/ff686783.aspx#section1 [online] 7.11.2011
[22] http://www.projecthosts.com/pol.asp
[23] www.unistav.sk
[24] The NIST Definition of Cloud Computing (Draft), Recommendations of the National Institute of Standards and Technology, Special Publication 800-145 (Draft), January 2011 http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf

Optimalizácia výroby pomocou procesného modelovania

Veronika VARGOVÁ, František BABIČ

Katedra kybernetiky a umelej inteligencie, Fakulta elektrotechniky a informatiky,
Technická univerzita v Košiciach, Slovenská republika

veronika.vargova@student.tuke.sk, frantisek.babic@tuke.sk

***Abstrakt* — Cieľom tohto článku je predstaviť prípadovú štúdiu implementácie procesného modelovania do výrobného procesu firmy HOWE Slovensko s.r.o., s cieľom optimalizácie používaných prístupov z pohľadu šetrenia nákladov a zníženia objemu nepodarkov. Výsledok predstavuje vytvorené procesné modely podľa štandardu BPMN, v rámci ktorých sú identifikované problémové miesta a na ich základe navrhnuté opatrenia na riešenie a zlepšenie. Tieto opatrenia umožnia zlepšiť organizáciu výrobných činností vo firme, zefektívniť výrobnú prevádzku a v neposlednom rade zvýšiť vyťaženosť vstupného materiálu.**

***Kľúčové slová* — Procesné riadenie, modelovanie, optimalizácia**

I. Úvod

Oblasť modelovania a analýzy procesných modelov sa stáva v mnohých firmách veľmi častým a dôležitým javom. Podnikové procesy sú jedným z najdôležitejších aktív podniku, ktoré umožnia firme byť krok pred konkurenciou. Takáto konkurenčná výhoda sa odráža najmä v schopnosti firmy rýchlo a pohotovo reagovať na zmeny trhu. Preto každá firma musí detailne poznať svoje procesy. Vďaka komplexnej analýze môže objaviť procesy, ktoré sú nemerateľné, nie sú zdokumentované alebo im pracovníci nerozumejú. Pomocou komplexnej analýzy je možné eliminovať nedokonalosti, ktoré sa objavujú pri vykonávaní stanovených firemných cieľov. Procesy, ktoré sú nemerateľné, nie sú zdokumentované alebo im pracovníci nerozumejú, sa po implementovaní procesného modelovania do prostredia výroby stávajú hodnotnými procesmi, ktoré majú presne stanovenú hodnotu prínosu.

Tento článok je rozdelený do štyroch hlavných častí, kde prvá čitateľa uvedie do problematiky procesného modelovania a riadenia. V druhej kapitole sú rozobraté problémy, ktoré boli identifikované na základe analýzy výrobného procesu kožených materiálov pre automobilový priemysel. Výsledky analýzy boli použité ako podklady pre návrh optimalizovaného procesného modelu, ktorý je popísaný v tretej časti článku. V závere sú zhrnuté použité prístupy a dosiahnuté výsledky.

A. Procesné riadenie

Hlavná definícia procesného riadenia spoločnosti hovorí, že je to činnosť vedúca k transformácií funkčne orientovanej organizácii na organizáciu procesného typu. Pri zavedení procesného riadenia do organizácie je cieľom zlepšiť a zefektívniť výrobné procesy. Všetky procesy vo firme sú podrobené analýze a cieľom je zistiť, ktoré procesy pracujú efektívne, a ktoré nevytvárajú výstupnú hodnotu. V konečnom dôsledku musí byť výrobný proces merateľný a neustále monitorovaný. Pri zobrazovaní procesov si musí organizácia uvedomiť, že tieto procesy nebudú slúžiť len jednotlivcovi, ale všetkým pracovníkom v organizácií, prípadne aj externým zložkám. Ak sa organizácia rozhodne zapojiť procesné modelovanie do svojej činnosti, je výhodné pracovať v takom modelovacom štandarde, ktorému bude každý rozumieť; kde bude presne dané, ako sa má tento štandard používať; prípadne aké grafické objekty sú v tejto norme štandardizované. Vytvorené procesné modely potom umožnia efektívne a kolaboratívne riadenie podnikových procesov.

Zavedením procesného riadenia firma získa detailné informácie o svojich podnikových procesoch, identifikujú sa silné a slabé stránky v procesoch, procesy budú merateľné a jednoznačne vyhodnotiteľné, a tiež sa znížia chyby vznikajúce pri ich vykonávaní .

B. Štandard BPMN

Pri modelovaní výrobných procesov bol použitý štandard BPMN, čiže Business Process Modeling Notation[1]. Tento štandard ponúka spoločnostiam možnosť porozumieť interným procesom v grafickej notácií tak, aby organizácie dokázali tieto procesy efektívne spravovať. Grafickú reprezentáciu procesov reprezentuje Business Process Diagram (BPD), ktorý tvorí sieť grafických elementov predstavujúcich aktivity a kontrolné body pre fungovanie procesu. Elementy, ktoré sa používajú pri modelovaní, sú rozpoznateľné od iných, a to z dôvodu využitia tvarov, ktoré sú špecifické pre modelovacie nástroje. Pomocou presne definovaných štandardizovaných objektov a možností ich použitia je vytvorený model zrozumiteľný pre každého používateľa.

II. PROCES VÝROBY

A. Popis firmy

Firma HOWE Slovensko s.r.o. sa zaoberá výrobou kožených materiálov pre automobilový priemysel prostredníctvom zastúpenia v rôznych krajinách sveta. Strihacia prevádzka, ktorá je otvorená v Košiciach, zameriava svoju činnosť na strihanie kožených dielcov na sedadlá do automobilov AUDI a Land Rover. Strihacie procesy pre jednotlivé typy automobilov sú od seba odlišné. Strihanie koží do automobilov AUDI je pomerne mladým procesom, a zároveň obsahuje viacero výrobných operácií. Firma neustále pracuje na zlepšovaní výrobných procesov a hlavným cieľom je minimalizovať počet kusov, ktoré neprešli schválenou kontrolou kvality a samozrejme dodržiavať termíny exportu dodávok zákazníkovi. V tomto procese existujú potenciálne možnosti na zlepšenie, preto bol podrobený analýze pomocou metód procesného modelovania.

B. Identifikované problémy v procese výroby

Hlavným problémom, s ktorým sa firma HOWE stretáva, je reklamácia zlých kusov. Výrobný proces AUDI je nastavený tak, že v prípade, ak zákazník dodané vyrobené kožené kusy považuje za nekvalitné, tak ich vráti späť a požaduje nápravu. Tento problém vzniká už pri výbere koží do skladu na začiatku výrobného procesu, kde neexistuje kontrola alebo procedúra, ktorá by triedila kože na základe požiadaviek. Tieto požiadavky určujú parametre, aké musí koža spĺňať pre rôzne typy vyrábaných materiálov (kusov).

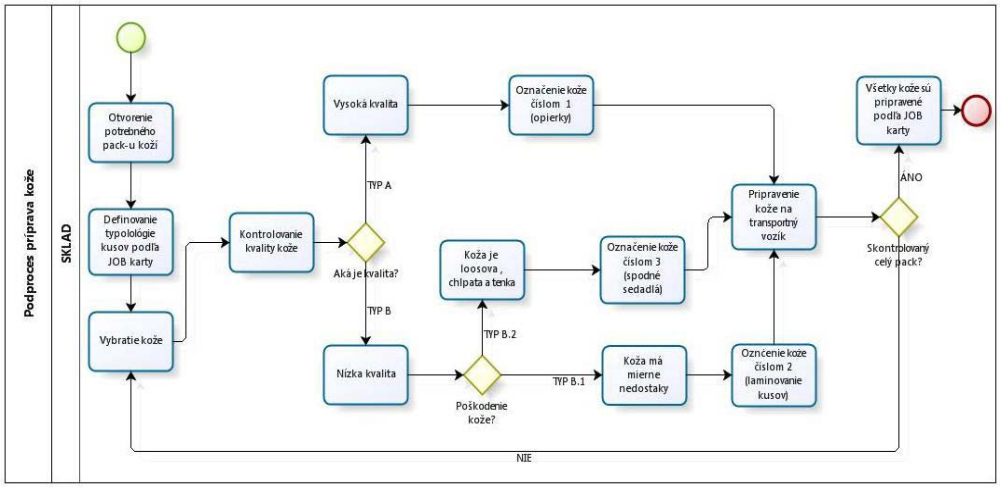
Druhý problém predstavuje vyťaženosť kože. Kožené kusy sa strihajú z celej kože, pričom základnou zásadou je, aby celková využiteľnosť kože bola čo najvyššia. Tento fakt ovplyvňuje skutočnosť, že na koži sú vyznačené zóny, do ktorých sa môžu ukladať len určité typy nožov. Zmeny pre dosiahnutie vyšších percentuálnych hodnôt vyťaženosti úzko súvisia so zmenami, ktoré budú aplikované v prvom prípade na zlepšenie kvality produkovaných kusov a tým pádom zníženie počtu reklamácií. Procesná zmena spôsobí, že kože budú rozdelené už v sklade podľa určenej typológie.

C. Optimalizácia výrobného procesu

Na základe analýzy bola identifikovaná možnosť vytvorenia nového podprocesu, ktorý by mal priamy vplyv na odstránenie identifikovaných problémov. Celý výrobný proces začína v sklade, preto bude tento nový podproces určený práve pre sklad. Obsah nového podprocesu s názvom „Sortovanie kože“ charakterizuje činnosť pracovníka na sklade, ktorý sa sústredí na dôkladnú kontrolu koží vstupujúcich ďalej do procesu. Tieto kože sú rozdeľované to troch kategórií. *Koža s číslom 1* spĺňa najvyššie kvalitatívne normy. Je určená na kusy, ktoré si vyžadujú najvyššiu kvalitu. V momente, keď sa táto koža dostane na strihací stôl, môžu byť na ňu umiestené nože, ktoré predstavujú najdôležitejšie kusy. *Koža s číslom 2* predstavuje laminovaciu kožu, ktorá sa použije na kusy určené na laminovanie, čiže kusy sa podlepujú špeciálnym materiálom. Vo všeobecnosti kožu charakterizuje nižšia kvalita, pretože laminovacie kusy sa podlepením kvalitatívne vylepšia. *Koža s číslom 3* sa vyznačuje loos-om, chlpatou rubovou stranou a hrúbkou je veľmi tenká. Takáto koža nemusí byť vyhodená, môže byť použitá na mini kusy a kusy, ktoré nie sú viditeľné v automobiloch. Pri tejto koži sa odrazí procesná zmena najviac, pretože sa pri strihaní obmedzí situácia, aby na kožu takého typu boli vystrihnuté kusy, ktoré si vyžadujú najvyššiu kvalitu.

Každá koža potrebná do zákazky bude prekontrolovaná pracovníkom, a ten ju následne zaradí do jednej z kategórií podľa kvalitatívnych noriem, viď Obr.1.

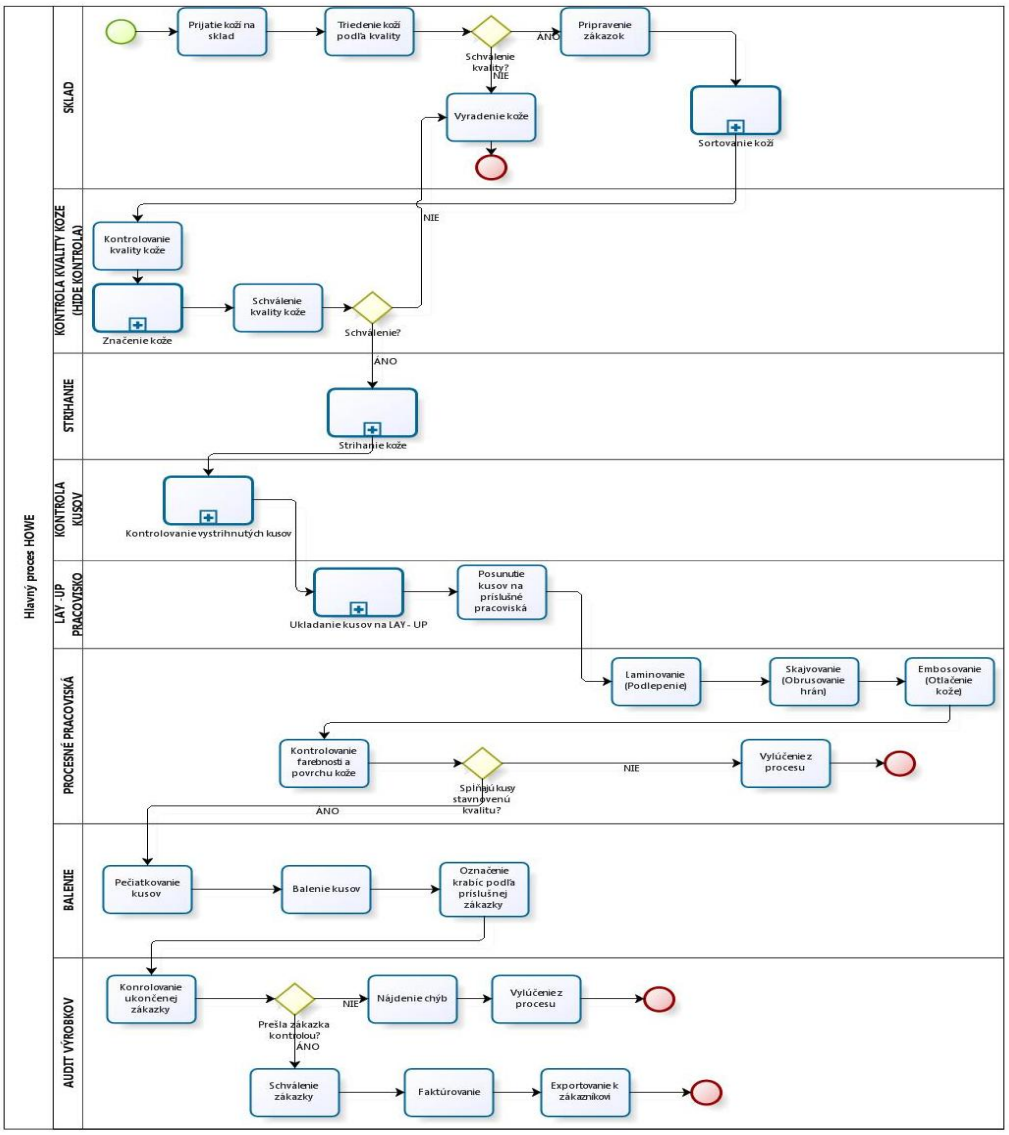
[1] http://www.bpmn-introduction.com



Obr. 1 Procesný model výberu správnej kože podľa podmienok zákazky.

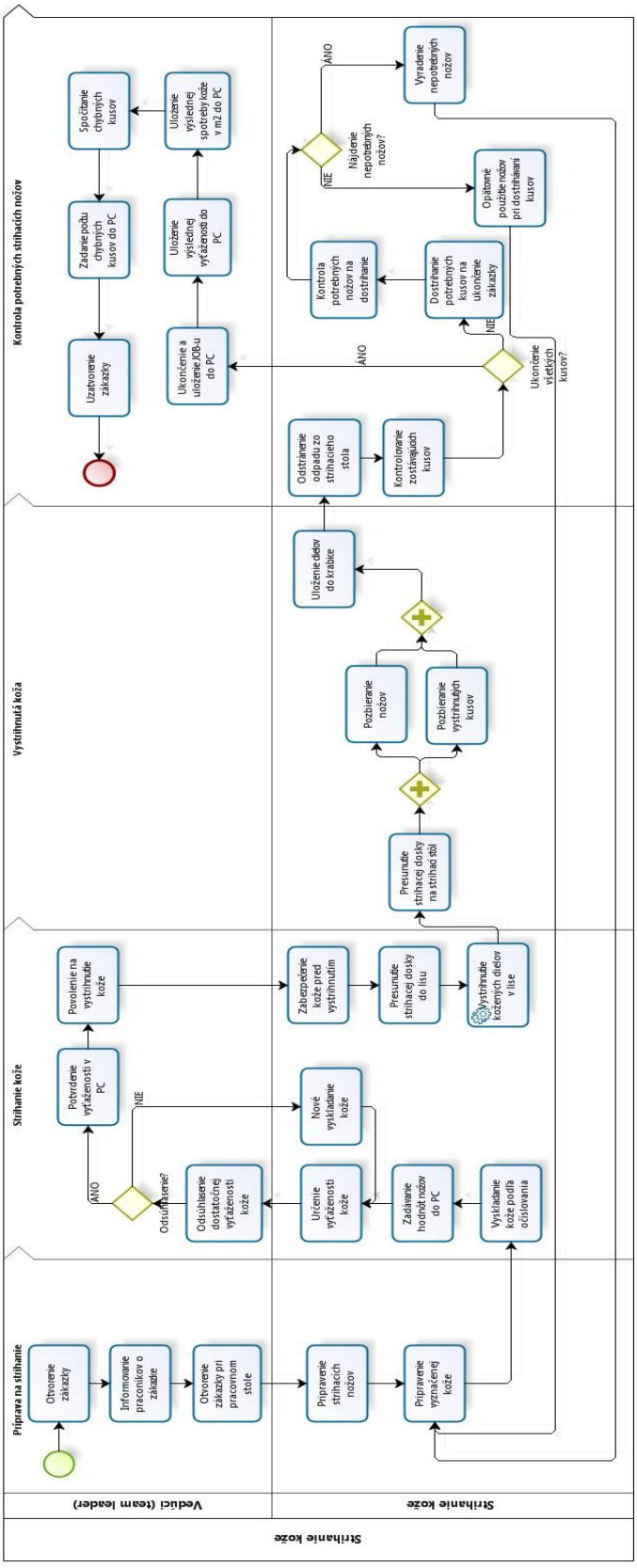
III. Optimalizovaný Procesný Model

Prezentovaná procesné zmena ovplyvní celý výrobný proces, ktorý je zobrazený na nasledujúcom obrázku.



Obr. 2 Hlavný výrobný proces po procesnej zmene.

Navrhnutá zmena ovplyvní aj proces strihania kože (viď. Obr.3), keďže pracovníci majú možnosť pracovať s kožou efektívnejšie. Pomocou rozdelenia koží vedia strihači prispôsobiť očíslovanej koži aj potrebné nože. Na kožu kat. 1 použijú konkrétne nože na vystrihnutie najkvalitnejších dielov; na kožu kat. 2 budú použité nože určené na laminovacie kusy a na posledný typ práve nože, ktoré nevyžadujú kvalitnú kožu.



Obr.3 Proces strihania po procesnej zmene.

IV. ZÁVER

Tento článok popisuje využitie metód procesného modelovania na optimalizáciu vybraného výrobného procesu na základe identifikovaných problémov. Pre tieto účely bolo vytvorených viacero procesných modelov, ktoré reprezentujú hlavný výrobný proces a jednotlivé podprocesy. Pre ilustráciu sú v tomto článku prezentované len ukážka hlavného výrobného procesu po navrhnutej procesnej zmene a model nového podprocesu pre triedenie kože. Tento doplnok pozmenil výzor hlavného modelu, a zároveň vylepšil fungovanie jednotlivých dotknutých podprocesov. Vytvorené modely popisujú každý výrobný detail tak, aby výsledok zodpovedal skutočnosti.

Cieľom nového podprocesu bolo zlepšiť vyťaženosť koží, ktoré vstupujú do procesu. Niektoré kusy si vyžadujú kožu špecifickej kvality, a práve táto skutočnosť je ošetrená pomocou navrhnutého opatrenia: triedenie koží v sklade podľa JOB karty. Táto zmena spôsobí, že každá koža bude zaradená do procesu s jasnou typológiou, čiže aké nože sa na ňu môžu položiť, a ktoré naopak nie, pretože kvalita kože je nedostačujúca. Toto opatrenie odstráni aj problém nedostatočnej kvality kože. Nový podproces upresňuje prácu na jednotlivých výrobných pracoviskách, ako sú: pracovné operácie v sklade, značenie kože, ale najmä strihanie kože.

Procesná zmena zároveň pozitívne ovplyvňuje aj povinnosti vedúcich pracovníkov. Títo pracovníci už nemusia sledovať proces prípravy koží v sklade, ale môžu svoju pozornosť venovať sledovaniu pracovníkov a zlepšovaniu výrobných prác.

Firma HOWE s.r.o. sa plne zaujímala o možnosti zlepšenia procesov a vytvorenie nového podprocesu hodnotila pozitívne. Po preštudovaní návrhu, vedúci kvality prijal tieto zmeny ako vhodný nástroj na zlepšenie výrobného procesu. Zavedenie takejto zmeny si samozrejme vyžaduje čas a finančné prostriedky na zaškolenie odborných pracovníkov, ktorí by sústredili svoju prácu len na odborné sledovanie a sortovanie koží. Keďže Firma HOWE neustále hľadá spôsoby zlepšenia svojich výrobných procesov, aj v tomto prípade má záujem o aplikáciu navrhnutej zmeny do svojej výrobnej činnosti.

POĎAKOVANIE

Tento príspevok vznikol za podpory KEGA projektu č. 065TUKE-4/2011 (50%) a realizáciou projektu Rozvoj Centra informačných a komunikačných technológií pre znalostné systémy (kód ITMS projektu: 26220120030) na základe podpory operačného programu Výskum a vývoj financovaného z Európskeho fondu regionálneho rozvoja (50%).

Evolučné prístupy v navigácii mobilných robotov

Ján VAŠČÁK

Centrum informačných a komunikačných technológií, Katedra kybernetiky a umelej inteligencie, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

jan.vascak@tuke.sk

***Abstrakt* – Tento dokument sa venuje návrhu systému s využitím evolučného prístupu pre navigáciu mobilného robota vo futbalovej hre „bago“. Hra je založená na prihrávaní lopty medzi hráčmi za účelom aby ju obranný hráč nechytil. Vzhľadom na to bol navrhnutý systém, ktorého základom je fuzzy regulátor, pričom požadované parametre toho regulátora sa nastavovali genetickým algoritmom. Pomocou genetického algoritmu sme upravovali bazu pravidiel regulátora. Systém bol navrhnutý v simulačnom prostredí *Webots robotics*, kde bol aj aplikovaný na humanoidného robota NAO pri hre bago.**

***Kľúčové slova* - Genetický algoritmus, Fuzzy regulátor, Nao.**

I. ÚVOD

Evolúcia ľudstva je tisíce až milióny rokov vývoja, kedy sa z najjednoduchších jednobunkových organizmov stávali zložité mnohobunkové živočíchy. Náš telesný vývoj už dosiahol hranice, ale naša inteligencia je ešte len na začiatku. V mojom programe vyvíjam mobilných robotov, ktorý sa snažia chápať, myslieť a vnímať to čo sa okolo nich deje. Reagovanie robotov síce nie je veľmi rýchle, ale tak ako sa vyvíjalo ľudstvo, vyvíja sa aj technika konštrukcie mobilných robotov. Pomocou evolučného prístupu som sa snažil naprogramovať robotov tak, aby sa aspoň trošku podobali skutočným futbalistom. Vybral som také metódy, spôsoby a postupy, aby robot vedel rozpoznávať svoju pozíciu, farbu a uhol svojho spoluhráča, uhol svojho protihráča a vzdialenosť a uhol lopty. Program by mal fungovať tak, aby robot vedel rozmýšľať, kde sa nachádza lopta a komu má loptu kopnúť a komu zasa nie. Evolúcia, čiže vývoj robotov ešte len začína a momentálne je na dobrej ceste za zdokonaľovaním.

II. NAVIGÁCIA MOBILNÉHO ROBOTA

Navigácia mobilných robotov v prostredí vyžaduje schopnosť robota dané prostredie určitým spôsobom vnímať. Na to slúži senzorový systém robota. Ten poskytuje aktuálne informácie o okolitom prostredí, ktoré spracúva riadiaci systém robota. Výstupom riadiaceho systému je informácia (napr. vzdialenosť prekážky, uhol od prekážky atď..), ktorá slúži na vykonanie potrebného kroku pri ceste z štartovacieho bodu do bodu cieľového. Veľká časť výskumných inžinierov sa špecializuje na vytvorenie takého algoritmu, pomocou ktorého by sa robot vedel orientovať v akomkoľvek neznámom prostredí. Podľa [276] navigácia je:

Je proces určovania a zachovania kurzu či trajektórie z jedného miesta na druhé. Procesy, ktoré odhaľujú pozíciu jednotlivca vzhľadom na poznaný svet, sú v navigácii kľúčové. Poznaný svet sa skladá z plôch, ktorých jednotlivé umiestnenia sú zobrazené na mape.

Navigácia sa z pohľadu snímania prostredia rozdeľuje do dvoch skupín na Globálnu a Lokálnu.

- *Globálna navigácia* je to navigácia robota pri ktorej je potrebné vedieť globálnu informáciu o prostredí. Taká navigácia sa vykonáva na základe poznania mapy priestoru, ktorú nám poskytujú kognitívne mapy. Kognitívna mapa sa vytvára napríklad z umiestnenia kamery nad prostredím.

- *Lokálna navigácia* je taká navigácia, pri ktorej sa určuje nasledovný krok len zo senzorických vlastnosti robota. Robot nemá informáciu o každom bode prostredia, len o určitej časti, v ktorej sa daný robot nachádza a pokiaľ mu to umožňujú senzorické vlastnosti.

III. INTELIGENTNÉ METODY NAVIGÁCIE MOBILNÉHO ROBOTA

Aplikácia metód a algoritmov umelej inteligencie pri riešení zložitých problémov rozhodovania a riadenia tak v technických, ako aj v netechnických vedách zaznamenala nebývalý rozvoj hlavne vďaka vzniku metód výpočtovej inteligencie (fuzzy množín, umelých neurónových sietí a evolučných algoritmov). Inteligentné systémy vyznačujúcimi sa aplikovanými prvkami umelej inteligencie sa sú charakterizované nasledujúcimi vlastnosťami:

- Majú schopnosť sa učiť a získavať znalosti
- Majú schopnosť dané znalosti uchovávať
- Vedia nadobudnuté znalosti použiť na riešenie problému

Podľa existujú úrovne riadenia inteligentných systémov, ktoré rozdeľujeme:

- Nultá úroveň - Robustný regulátor
- Prvá úroveň – Adaptívny regulátor
- Druhá úroveň – Optimálny regulátor
- Tretia úroveň - Plánovací regulátor

Robustný regulátor: Je to typ regulátora, ktorý je robustný voči zmenám respektíve zachovávajú sa jeho kvalitatívne vlastnosti (stabilita, kvalita regulácie). Slovom *Robustnosť* môžeme definovať systém, ktorý sa správa voči zmenám a poruchám algoritmu neohľaduplne respektíve jeho ďalšie pôsobenie sa zaobíde len s malou ujmou na spôsobe regulácie.

Adaptívny regulátor: Používa sa v prípadoch kde sa v priebehu času menia dynamické alebo statické vlastnosti regulovanej sústavy (napr. zmena koeficientu prestupu tepla usadeninami, zmena prietoku zanášaním atď.), ktoré môžu mať za následok zhoršenie kvality regulácie, nakoniec i jej zlyhanie. Adaptívny regulátor je schopný tieto zmeny identifikovať a prestaviť svoje konštanty, tak aby bola zaistená pôvodná kvalita regulácie.

Optimálny regulátor: Je regulátor, ktorý má z určitého hľadiska najlepšie možné vlastnosti. Môžeme napr. požadovať odstránenie regulačnej odchýlky v minimálnom čase, alebo s najmenším preregulovaním (dynamická optimalizácia), alebo s vynaložením minimálnej energie alebo s maximálnou účinnosťou.

Plánovací regulátor: Je regulátor, ktorého obsah pozostáva z plánovacích funkcii, ktorých náplňou je:

- Plánovanie rizika
- Plánovanie porúch
- Plánovanie chýb
- a pod....

IV. GENETICKÝ ALGORITMUS

Prírodná evolúcia je úspešná a robustná adaptácia v biologických systémov.

Genetický algoritmus sa objavili začiatkom 60. rokov minulého storočia a nachádzajú sa na pomedzí deterministických a nedeterministických algoritmov. Princíp práce genetického algoritmu je postupná tvorba generácii rôznych riešení daného problému. Pri riešení sa uchováva tzv. populácia ktorej každý jedinec predstavuje jedno riešenie daného problému. Ako populácia prebieha evolúciou riešenie sa zlepšuje.

Jedince sa označujú názvom fenotyp a pre jeho reprezentáciu sa používa názov genotyp, alebo chromozóm. Chromozóm sa delí na jednotlivé gény, ktoré sú lineárne usporiadane. Gén nadobúda rôzne hodnoty. Jedinci môžu byť zakódované rôznym spôsobom. Spôsob ako je jedinec zakódovaný ma veľký vplyv na úspech alebo neúspech riešenia. Najjednoduchším spôsobom ako zakódovať jedinca je binárny jedinec nejakej dĺžky. (2)

(1,0,1,0,1,1,0,0) (1)

1. Inicializácia – Vytvorenie prvotnej populácie
2. Začiatok – Na základe určitej metódy vyberieme z populácie niekoľko jedincov s vysokou zdatnosťou
3. Pomocou genetických operátorov vytvoríme novú populácie
   - Kríženie – Výmena časti jedincov medzi sebou
   - Mutácia – Náhodná zmena časti jedinca
   - Reprodukcia – Kopírovanie jedinca bez zmeny
4. Vyhodnotenie jedincov novej populácie
5. Ukončenie algoritmu na základe ukončovacej podmienky
6. Koniec algoritmu – Výstupom je jedinec s najlepšími vlastnosťami

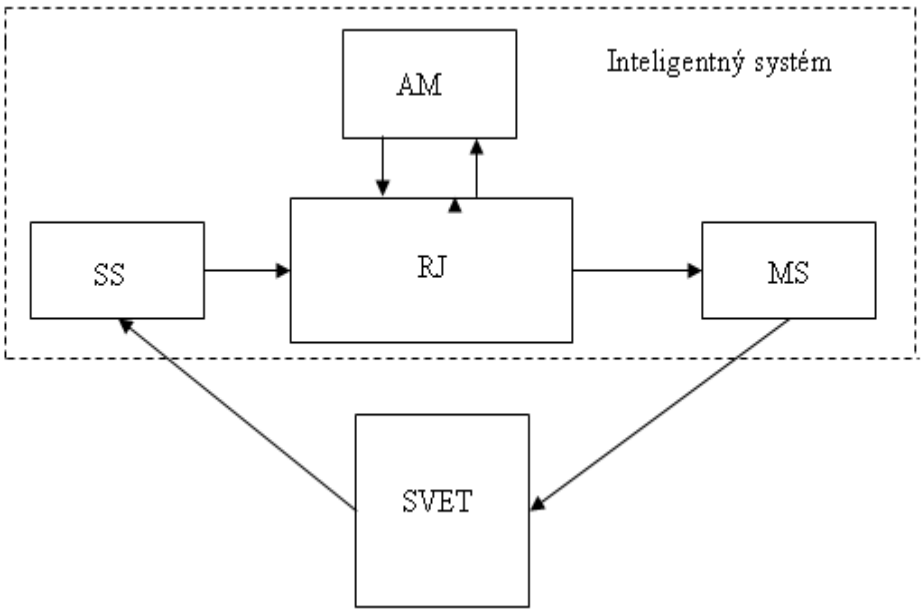
Genetické algoritmy sú väčšinou používané v praxi na optimalizáciu úloh, ale miernou úpravou toho algoritmu ho môžeme použiť na rozmanité úlohy.

V. NÁVRH SYSTÉMU

Štruktúra systému pozostáva zo štyroch časti, ktoré sú navzájom prepojené:

1. Senzorický systém (SS)
2. Riadiaca jednotka (RJ)
3. Motorický systém (MS)
4. Adaptačný mechanizmus (AM)

Štruktúra systému je znázornená na obrázku (Obr. 1)



Obr. 1 Štruktúra inteligentného systému

A. *Senzorický systém*

Na zber dát z prostredia bola použitá kamera, ktorá plnila účel, ako sú u ľudí oči a na základe klasických goniometrických funkci sa dané vzdialenosti a uhly objektov dokázali prepočítavať

B. *Riadiaca jednotka*

Riadiaca jednotka tvorila jadro navigácie robota a pozostávala s fuzzy regulátora typu TSK (Takagi – Sugeno - Kang) je to typ regulátora, ktorý nemá schopnosť adaptácie a preto som tam vytvoril hybridný prístup, kde na učenie fuzzy regulátora bol použitý genetický algoritmus.

Fuzzy regulátor pozostáva s piatich časti: (Obr. 2)

1. Fuzzyfikácia – Mení ostré hodnoty na stupne príslušnosti jednotlivých lingvistických premenných.
2. Báza pravidiel – Obsahuje jednotlivé pravidla typu IF – THEN
3. Báza dát – Definuje jednotlivé funkcie príslušnosti
4. Vyhodnotenie pravidiel - je určenie výsledného stupňa príslušnosti celej predpokladovej časti pre každé pravidlo.
5. Defuzzyfikácia - je prevod výslednej funkcie príslušnosti výstupov na reálnu hodnotu (ostrú).

Fuzzy regulátor typu TSK je o niečo jednoduchší. Pri tomto type regulátora odpadá potreba defuzzyfikácie, čo ma za následok to, že obsahuje pravidla v tvare: (2)

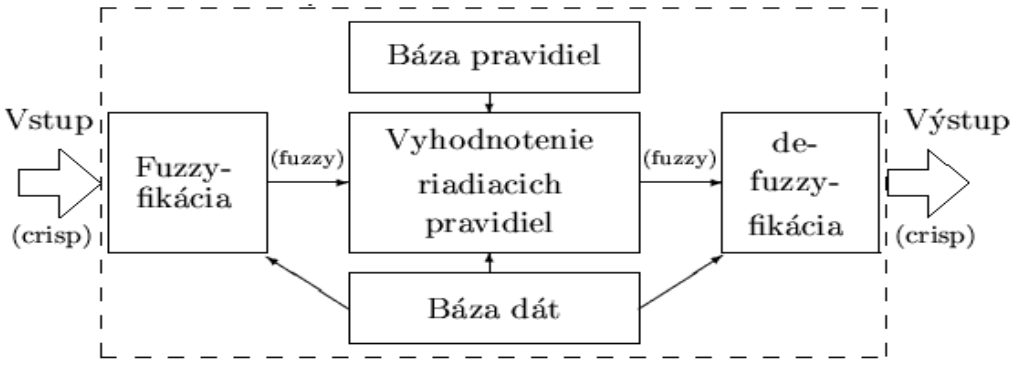
$$\text{Ak } x \text{ je } A \ \& \ y \text{ je } B \text{ Potom } z = f(x,y) \qquad (2)$$

Kde:

x , y – sú vstupy privádzané do fuzzy regulátora.

A , B – sú funkcie príslušnosti priradené danému vstupu.

z – je funkcia vstupu x a y, už ostrá hodnota jedného pravidla.



Obr. 2 Štruktúra Mamdanýho fuzzy regulátora

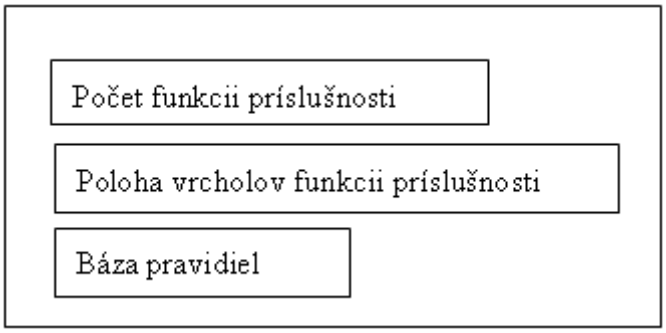
C. Adaptačný mechanizmus

Samotný TSK fuzzy regulátor je neadaptívny typ regulátora a tak máme snahu tento fakt zmeniť a vytvárame hybridný prístup, ktorý rozširuje daný TSK regulátor o schopnosť vyvíjať sa. Ako adaptačný mechanizmus sme použili Genetický algoritmus, ktorý má na základe výskumu najlepšie vlastnosti v oblasti riadenia. Genetický algoritmus patrí do skupiny evolučných algoritmov. Samotná adaptácia fuzzy regulátora sa týka úlohy zmeny bázy pravidiel, kde sa populácia skladá z jedincov, kedy jeden jedinec predstavuje jedno riešenie celého problému.

D. *Štruktúra jedinca*

Samotná štruktúra jedinca pozostáva z parametrov fuzzy regulátora: (Obr. 3)

1. Počet funkcií príslušnosti
2. Poloha vrcholov funkcií príslušnosti
3. Samotná báza pravidiel



Obr. 3 Štruktúra jedinca

Prvá časť chromozómu pozostáva z trojprvkového chromozómu odpovedajúcemu počtu vstupov do regulátora. Pre každý vstup je vygenerovaný počet funkcii príslušnosti z intervalu <3,7>. Táto časť chromozómu sa počas adaptácie po vygenerovaní nemení.

Druhá časť chromozómu obsahuje hodnoty vrcholov funkcii príslušnosti. Keďže na základe vrcholov dokážeme určiť polohu jedného a polohu druhého nosiča funkcie príslušnosti potrebného na výpočet stupňa príslušnosti k danému vstupu. Dĺžka tohto chromozómu je rôzna a závisí od počtu nadefinovaných funkcii príslušnosti. Pri tejto časti chromozómu je dôležité zachovať fuzzy partíciu. Táto čas chromozómu sa počas adaptácie po vygenerovaní nemení.

Posledná časť chromozómu je venovaná tvorbe pravidiel, jeho počet je počas adaptácie premenlivý, kde jedno pravidlo predstavuje päť prvkové pole (

Obr. 4). Počet vygenerovaných pravidiel je zo spodnej hranice ohraničený na zvládnutie danej úlohy a hornú hranicu ma stanovenú na základe počtu funkcii príslušnosti vygenerovaných pre dané vstupy do regulátora. Maximálny počet pravidiel pre jedného jedinca je vypočítavaný na základe $N_{I1} x N_{I2} x N_{I3}$.

Kde N_{I1} - predstavuje počet funkcii príslušnosti pre prvý vstup.

1	3	7	0.15	0.5

Obr. 4 Štruktúra pravidla

- Prvé tri pozície pravidla prislúchajú vstupom a hovoria ktorá funkcia príslušnosti je priradená danému vstupu.
- Štvrtá pozícia je priradená hodnota výstupu z intervalu <-1.57;1.57>.
- Piata pozícia v pravidle prislúcha váhe pravidla nadobúda hodnoty z intervalu <0;1>.

VI. GENETICKÝ ALGORITMUS PRE NÁVRH BAZY ZNALOSTI

Populáciu tvorí určitý počet jedincov, kde jedinec predstavuje jeden návrh fuzzy regulátora. Štruktúra jedinca pozostáva z niekoľko typov chromozómov. Úlohou genetického algoritmu je navrhnúť takú bazu znalosti, ktorá by mala najlepšie vlastnosti na riešenie daného problému.

Vytvorenie populácie

Vytvorenie populácie je proces pri ktorom sa jedinci nagenerujú náhodne a to podľa vopred definovaných podmienok .

Ohodnotenie jedincov

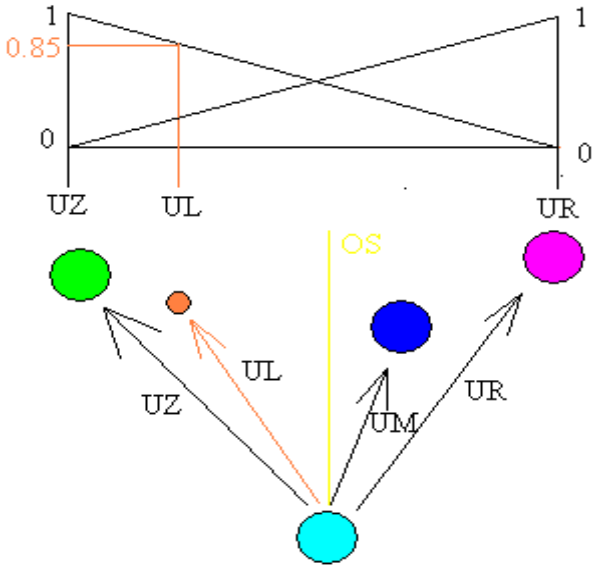
Každý jedinec po vytvorení nesie zo sebou svoje vlastnosti či už dobre alebo zle. Aby sme vedeli či vlastnosti daného jedinca sú vhodné pri vytváraní správania potrebného na zvládnutie úlohy je potrebné ich nejakým spôsobom merať. Pri meraní schopnosti jedinca použijeme fitness funkciu, ktorá bola definovaná na základe presnosti nahrávky daného hráča

svojmu spoluhráčovi. Grafické znázornenie prepočtu funkcie príslušnosti je na obrázku (Obr. 5)

Na určovanie hodnoty funkcie príslušnosti som použil jednoduchý trojuholník, kde som prepočítaval hodnoty na základe analytickej geometrie podľa vzorca (3)

$$fitness = \frac{UholH2 - UholH1}{UholH2 - UholLopty} \tag{3}$$

Takýto spôsob sa používal na ohodnotenie jednej hracej situácie. Pri zadefinovaní väčšieho počtu hracích kombinácii sa fitness prepočítala ako aritmetický priemer cez všetky hracie kombinácie čo nám umožnilo, že daný jedinec nezabudne naučenú kombináciu pri učení inej, ale sa kontroluje správanie jedinca pri viacerých možných hracích kombinácii a vypočítaná hodnota fitness popisovala vlastnosti daného jedinca.



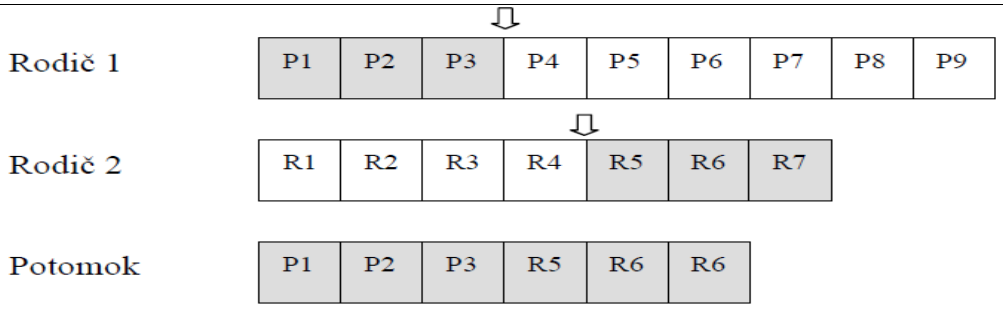
Obr. 5 Výpočet hodnoty fitnes funkcie

Selekcia

Tento blok genetického algoritmu je veľmi dôležitý pre tvorbu jedincov do novej populácie. Podstata selekcie je výber dvoch rodičov, ktorý sa budú podieľať na reprodukčnom procese, kde z dvoch vybraných rodičov sa vytvorí jeden potomok. Pri selekcii rodičov sa museli zadefinovať ohraničenie kvôli možnému kríženiu rodičov. Ohraničenie spočívalo v tom, že do reprodukčného procesu sa mohli dostať len taký rodičia, ktorý majú rovnakú časť štruktúry chromozómu, v podstate sa jedná o to, že musia mať rovnakú prvú časť chromozómu, čiže rovnaký počet nadefinovaných funkcii príslušnosti pre dané vstupy. Rodičia bolí vyberaný náhodným spôsobom či už obsahovali dobré alebo zle vlastnosti, čo nám prinieslo väčšie množstvo genetického materiálu s ktorým mohol daný genetický algoritmus pracovať.

Kríženie

Pri tomto procese som použil metódu jednobodového kríženia. Podstata spočíva v tom, že rodičia môžu mať rôzny počet nadefinovaných pravidiel a pre každého z rodičov sa náhodne vygeneroval rôzny bod kríženia.(Obr. 6) Po úspešnom vytvorení potomka sa musela uplatniť jedna modifikácia jedinca, ktorá odstránila potenciálne rovnaké pravidla. Takýto jedinec bol vložený do populácie potomkov. Celý proces sa opakoval pokiaľ nebola vytvorených toľko potomkov koľko bolo rodičov.



Obr. 6 Jednobodové kríženie rodičov

Vytvorenie novej populácie

Nová generácia bola vytvorená výberom medzi populáciou rodičov a populáciou potomkov, na základe dosiahnutých vlastnosti pri riešení daného problému. Dosiahnuté vlastnosti sa ohodnocovali fitness funkciou. Pri vytvorení novej populácie bol použitý mierny eltizmus, kde jedinci s najlepšími vlastnosťami s predošlej generácie boli hneď presunutý do novej generácie, pričom týmto prechodom nestratili možnosť reprodukcie.

VII. EXPERIMENTY

Pri vývoji môjho genetického algoritmu som sa rozhodol nastavovať parametre, ktoré priamo súvisia z algoritmom a ktoré súvisia len okrajovo respektíve súvisia, ale s modifikáciou genetického algoritmu. Nastavované parametre sú:

1. Počet jedincov
2. Počet generácii
3. Počet herných kombinácii
4. Medzná hodnota podobných pravidiel
5. Počet elitických jedincov

Zmenou tých to pravidiel by som chcel zlepšiť evolučný proces a tým dosiahnuť uspokojivé výsledky pri navigácii mobilného robota v futbalovej hre „bago“.

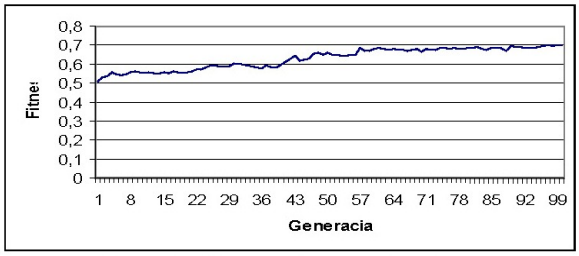
Počet jedincov ma veľký vplyv na riešenú úlohu pričom na istý počet herných kombinácií je prislúcha úmerný počet jedincov. Väčšie množstvo herných kombinácii si vyžaduje veľké množstvo znalostí, ktoré sú uložené v genetickom materiálu jedincov. Zväčšenie genetického materiálu dosiahnem zväčšením počtu jedincov čiže veľkosti populácie. Toto správanie si môžeme všimnúť na experimentoch 1 - 4. Podľa toho vyplýva že na 5 herných kombinácii potrebujeme aspoň 15 krát väčšie množstvo jedincov.

Pri zmene prahovej hodnoty podobných pravidiel nastáva problematická otázka, kedy sú dane pravidla rovnaké a kedy umelo ovplyvňuje evolučný proces tým že zatracujeme genetický materiál, ktorý by mohol byť pre nás dosť zaujímavý. Tento návrh som robil metódou pokus omyl. Tato metóda nebola až tak veľmi nepriateľna lebo ak sme nastavili prahovú hodnotu vyššiu tak genetický algoritmus nám neprinášal tak dobré výsledky a naopak ak hodnota prahu bola príliš nízka začali sa nám hromadiť pravidla čo malo za následok, že pomalý sa nám genetický materiál začal dosť na seba podobať a jedinec ostal v lokálnom minime. Pri ideálnom nastavení sa prahovej hodnoty sa výrazne zmenila najlepšia vhodnosť jedinca.

VIII. ZÁVER

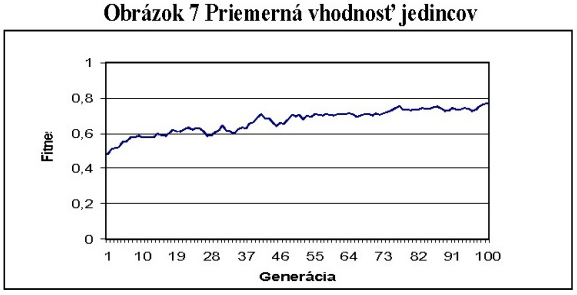
Použitie fuzzy regulátora s adaptáciou na riadenie smeru nahrávky pri futbalovej hre (bago) sa javy ako veľmi dobra voľba. Fuzzy regulátor sa mal adaptovať na 30 herných kombinácii čo si vyžaduje veľký počet uchovávaných znalosti. Veľkou výhodou použitia fuzzy regulátora s možnosťou adaptácie je, že ak na vstup príde podobná situácia ako je situácia na ktorú bol naučený dokáže veľmi pekne reagovať a výsledky sú pomerné dobré. Pri vyvinutí takých základných herných kombinácii, ktoré by pokrývali celý herný priestor, dokázal by robot s použitím fuzzy regulátora s možnosťou adaptácie reagovať na akúkoľvek situáciu a priniesť kvalitný výsledok, či už na ňu naučený je alebo nie.

Počet jedincov: 50.
Počet herných kombinácii 10.
Počet eliťákoch: 5.
Počet generácii: 100.



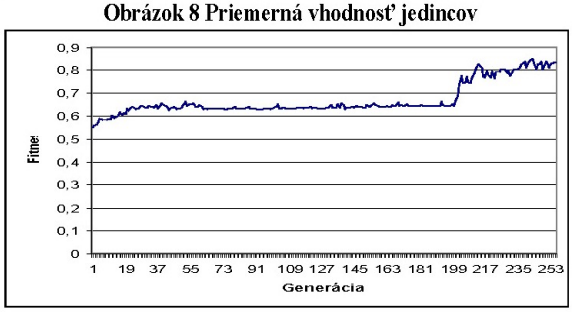
Obrázok 7 Priemerná vhodnosť jedincov

Počet jedincov: 50.
Počet herných kombinácii 5.
Počet eliťákoch: 5.
Počet generácii: 100.



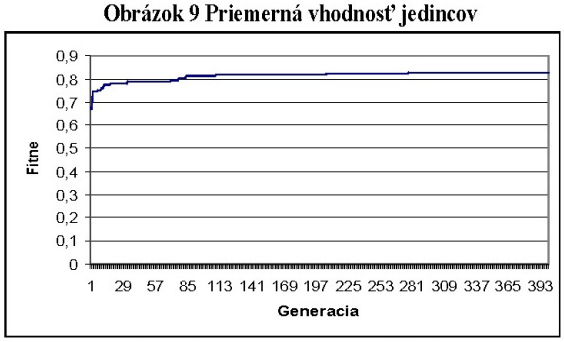
Obrázok 8 Priemerná vhodnosť jedincov

Počet jedincov: 100.
Počet herných kombinácii 5.
Počet eliťákoch: 15.
Počet generácii: 253.



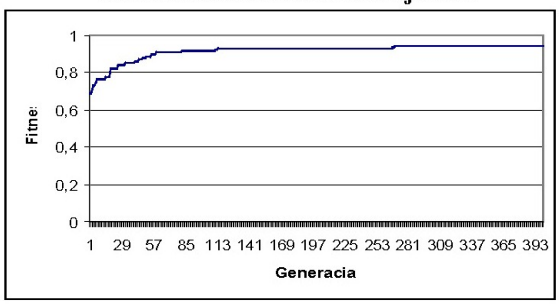
Obrázok 9 Priemerná vhodnosť jedincov

Počet jedincov: 300.
Počet herných kombinácii 15.
Počet eliťákoch: 15.
Počet generácii: 400.
Medzná hodnota: 0.25



Obrázok 10 Priemerná vhodnosť jedincov

Počet jedincov: 300.
Počet herných kombinácii 15.
Počet eliťákoch: 15.
Počet generácii: 400.
Medzná hodnota: 0.15



Obrázok 11 Priemerná vhodnosť jedincov

POĎAKOVANIE

Táto práca bola vytvorená realizáciou projektu Rozvoj Centra informačných a komunikačných technológií pre znalostné systémy (kód ITMS projektu: 26220120030) na základe podpory operačného programu Výskum a vývoj financovaného z Európskeho fondu regionálneho rozvoja.

LITERATÚRA

[1] JURIŠICA, L: LAT&T Journal 5/2003 – Robotika

[2] NASCIMENTO, J: Artificial NN in Control and Optimization. Doctor thesis, Manchester UK: Faculty of Technology, University of Manchester 1994.

[3] SINČAK P. a ANDREJKOVÁ G.: Neurónové siete Inžiniersky prístup 1. Elfa: Kosice, 1996.

[4] Vaščak J.: Fuzzy logika v regulácii. Podrobný úvod do problematiky fuzzy riadenia. 2008. Dostupné na internete:

<http://www.ai-cit.sk/source/fuzzy/FLvR.pdf>

[5] I. Sekaj: Evolučné výpočty, IRIS Bratislava, 2005

Use of Multi-agent Approaches in Robotic Soccer

Ján VAŠČÁK

Center for Intelligent Technologies, Department of Cybernetics and Artificial Intelligence, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Slovak Republic

jan.vascak@tuke.sk

***Abstract*— In this paper introduction into robotic soccer strategies based on multi-agent approach is presented. Research on current state of art of methods and policies, used in play strategy in different leagues (FIRA, RoboCup etc.) is conducted. Selected method is applied in domain of robotic soccer according to the rules of SPL league of RoboCup initiative. We evaluate this method applicability on model of Nao robot, produced by Aldebaran Robotics, in simulation environment Webots.**

***Keywords*—Nao, robotic soccer, multi-agent systems, RoboCup SPL**

I. INTRODUCTION

Term multi-agent systems (MAS) overlays complex set of methods and tasks, which main goal is to create coherent system, composed of autonomous units (agents) capable of achieve desired goal, using only partial knowledge and limited abilities [1]. Multi-agent systems were inspired in economics, sociology and biology. They were developed as intersection of distributed computing and artificial intelligence under the term distributed artificial intelligence. Main idea of distributed computing is possibility to divide complex computing problem into several process and therefore, distributed artificial intelligence vary from distributed computing in way of problems solving thru communicating and cooperation compared to low-level synchronization and parallelization [2]. Basic unit of MAS is Agent. Though term Agent has no exact definition, in literature general agent is referred as following [3]: pro-activity, reactivity, autonomous behavior and social behavior. Separating agents into different groups based on their own specific abilities and methodologies of agent building we can divide MAS into several architectures: Deliberative architecture (first think, then act), reactive architecture (do not think, act), hybrid architecture (think and act simultaneously, parallel), behavior/ based architecture (think about way, how to act). Features characteristic for MAS are: computing is asynchronous, knowledge is decentralized, no control on global level, each agent has only partial information about problem or insufficient abilities to solve problem. MAS are classified by two features homogeneity and centralism. Homogenous MAS contains only homogenous agents-, agents with identical or similar features and abilities. Heterogeneous MAS includes agents with different abilities and each agent or group of agents has own specialized goal to achieve. Communication between agents is a must in this system. If we consider centralism as criteria we can divide MAS into centralized (system based on only one agent) and decentralized (multiple agent) systems. Very important feature of MAS is ability to cooperate and communicate between agents. Two common methods of communication are used- direct and indirect communication. Direct communication is broadcasting as sending/receiving information to all agents at time, multicasting as sending/receiving messages between specified group of agents and aimed information passing to certain agent. Indirect communication is based on idea of shared data structure often referred as table, which is available to all agents. Cooperation is considered as a main difference between MAS and related disciplines. Its aim is to boost problem solving with better parallel approach, lower concurrence between agents and to increase set of solvable problems.

ISBN 978-80-553-0890-6 © 2012 FEI TUKE

II. STATE OF ART IN ROBOTIC SOCCER DOMAIN

Robotic soccer is one of most interesting areas to test MAS, because it includes most of the problems that MAS ought to solve. Robotic soccer is well-known for its complexity, high degree of indetermination and for its dynamically changing environment.

A. Robot soccer in SPL RoboCup

Standard Platform League (SPL) of RoboCup initiative is a robot soccer league, which use Nao biped robot produced by Aldebaran Robotics as standard testing platform. RoboCup is organization dedicated to developing a team of fully autonomous humanoid robots that can win against the human world soccer team by year 2050. Soccer game in SPL is between two teams of Naos, 3 robots in each team. Official playground for league is on figure 1.

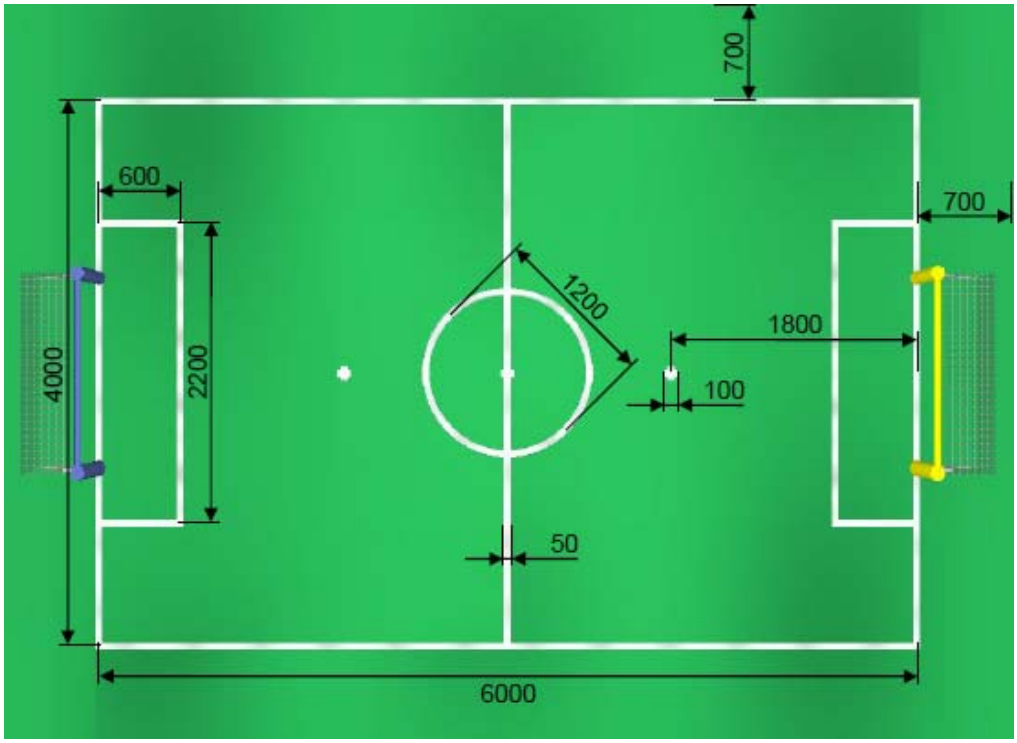


Fig. 1 SPL official playground

Except specific field some special rules also apply. Red team defend yellow goal, blue team defend blue goal. One player has to be designed as goalkeeper and cannot move from his penalty area; also it's the only player who can touch the ball with his hands. It wears number one. Other two players take role of field players. They cannot enter their penalty area, wear number two and three and in case they are penalized they are removed for 30 sec from field. Communication is allowed only between robots and GameController (kind of virtual referee), so robots play without human control. There is no restriction on communication between robots using a microphone or a speaker. Robots also can use wireless communication up to 500Kbps of bandwidth. Goal is achieved when the entire ball goes over the goal-side edge of the goal line. [4]

B. Webots

Webots is simulation environment, developed for modeling, programming and simulating mobile robots, from Cyberbotics. In this environment user is able to design a test complex robotic systems, with one or more heterogeneous and homogenous robots. For purpose of this paper we introduce only Robotstadium. Robotstadium is on-line competition in modeling robotic soccer with almost realistic model of Nao robot, under the rules of SPL. It supports programming languages Java, Python, C, and C++. Difference between real Nao and its model is in his degrees of freedom – 25 on real one compared to 22 on simulation model. [5]

C. Image processing in Webots

Model uses one camera with resolution 160*120 pixels. To cope with problem of two

cameras on real Nao this camera has two positions upper and lower with 40 degrees angle between them. Upper position capture snaps from horizon, lower position is for area under Naos feet. The problem is that between switching sometimes Nao gets into some sort of "blind" zone. It's a small angle between camera switches, where Nao is not able to catch picture as shown on figure 2. Area market as 2 is sight of view from upper position and area market 1 from lower position. Picture from camera is represented as array. Each object is represented as single pixel, for example ball is orange pixel. When searching for ball we simply use RGB color combination of orange color. Because Webots is able to simulate light conditions we have to add a threshold tolerance for color. In our case it will by R=240, G=140, B=50 and threshold=60. Similarly we are able to locate blue goalpost (30, 200, 200, and 60) and yellow goalpost (140, 140, 15, and 60). Refresh rate of camera is 0.16 sec.

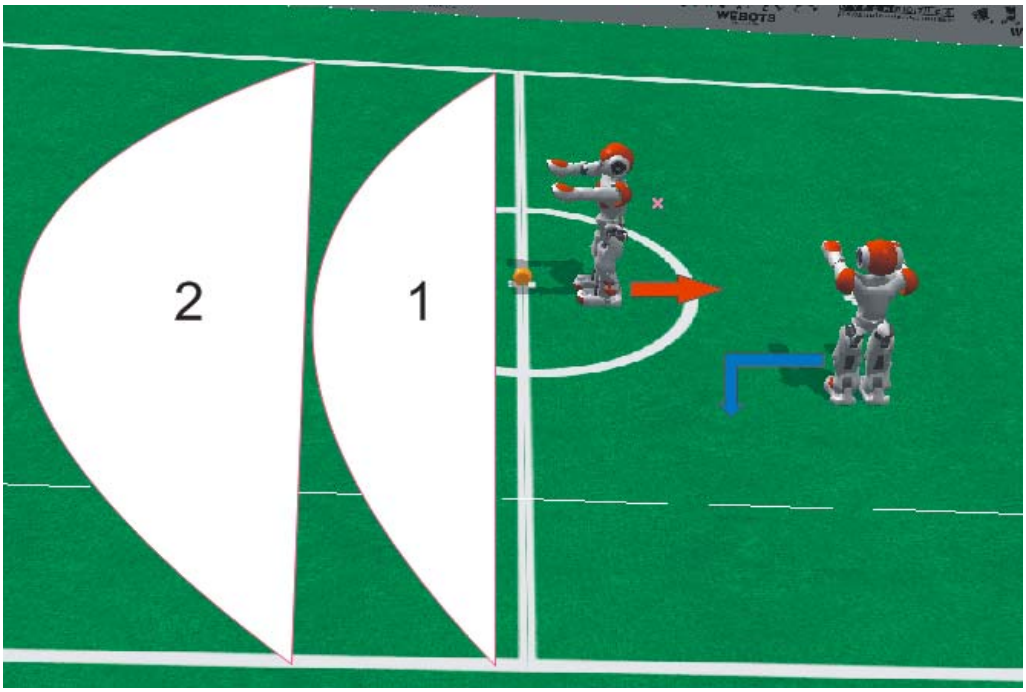


Fig. 2 Blind zone between camera positions

For acquiring distance between robot and goal or ball we can use height of camera, which is in 51 cm from ground and angle between head position and ball pixel – angle of head joint. Also we can estimate angle between Naos direct position (where Naos feet point) and horizontal turn of his head joint.

III. SELECTED METHODS AND STRATEGIES IN OTHER ROBOT SOCCER LEAGUES

In this section we introduce some strategies that are currently used by teams in other leagues than SPL and one from SPL, even if the team doesn't use MAS approach.

A. CMDragons

CMDragons is a team from the Carnegie Mellon University and competes in league of small robots. In each team play 5 robots of their own design. Overhead camera for scanning field and external computer for image processing is allowed. Their perception is secured by two overhead cameras and image processing by computer that estimate current position of ball and players of both teams. They estimate velocity and track with enhanced Kalman filter. Then they process information in module for strategy and tactics, which is composed from three layers. First one is for world state recognition, to decide if the team is in defensive or offensive. Next layer is responsible for behavior implementation from basic moves as "go to point" to more complex moves as pass or shot. Third layer is for navigation and path planning. They developed their own architecture called STP (skill-tactics-play) to cope with problems of dynamic role assignment. In this architecture pre-defined games are used to determine current goal for robot. They also use potential fields for obstacle avoidance – navigation. STP divides goals on "active" and "support" and for solving them uses different methods. More details on their tactics are in [6].

B. Northern Bites

Northern Bites is team from Bowdoin College and they compete in SPL league from 2006 on Aibo robodogs and Nao bipeds and in year 2009 they became a vice-champion in SPL. For position estimation they use extended Kalman filter. As reference point they use goals and corners specially angles and distance between them. One of the most challenging tasks is problem of data association. For accurate assignment of observed corner to real corner they use method of clustering – nearest neighbors. They model position of robot as well as position of ball, with its estimated dynamics. It's because ball has a kind of central position in soccer and this approach reduces computational time of world states. To improve movement they applied Zero Moment point with controller for trajectory generation. Their behavior modeling uses FSA – Finite State Automaton. To make complex problems simpler, they use multiple FSA, three for behavior controlling and one safety controller used in case of fall. They use eight formations, which depends on ball position. Also they assign four different roles to robot – goalkeeper, attacker, defender, and runner. Runner is player, which is closest to ball. It's up to each robot to decide who will be runner to reduce waiting time. Other roles are assigned depending on current formation. [7]

C. Marker-Driven Approach

Marked-Driven methodology is base on idea of free market, so consequently maximization of the overall system profit is a goal. Each participant in the market tries to maximize its profit, as a result of this; the overall profit for the system is expected to rise. The idea for multi-robot teams is based on the interaction of the robots among themselves in a distributed fashion for trading work, power and information. In general, there is a main goal of team. Some entity outside the team is assumed to offer payoff for that goal. The main goal of the system is decomposed into smaller tasks and auction is preformed for each of these tasks. In each auction robots communicate among themselves and calculate their estimated cost for accomplishing that task and offer a price to the auctioneer.At the end of the auction, the bidder with the lowest offered price will be given the right of execution of the task and receives its revenue on behalf of the auctioneer. There are many possible actions that can be taken. A robot may open another auction for selling a task that it won from another auction, two or more robots may cooperatively work and get a task which is hard to accomplish by a single robot, or for a heterogeneous system, robots with different sensors/actuators may cooperate by resource sharing (for example, a small robot with a camera may guide a large robot without a vision system for carrying a heavy load). It's possible to calculate each robot bid for auction as following.

$$C_{es} = \mu_1 t_{dist} + \mu_2 . t_{align} + \mu_3 . clear_{goal} \quad (1)$$

$$C_{bidder} = \mu_4 t_{dist} + \mu_5 . t_{align} + \mu_6 . clear_{teammate(i)} + C_{ES(i)}, i \neq robotid \quad (2)$$

$$C_{auctionerr} = C_{ES(robotiid)} \quad (3)$$

$$C_{defender} = \mu_7 t_{dist} + \mu_8 . t_{align} + \mu_9 . clear_{defense} \quad (4)$$

where *robotid* is the id of the robot, *tdist* is the time required to move for specified distance, *talign* is the time required to align for specified amount, μi are the weights of several parameters to emphasize their relative importance in the total cost function, *cleargoal* is the clearance from the robot to goal area, *clearball* is the clearance from the robot to ball, *cleardefence* is the clearance from the robot to the middle point on the line between the own goal and the ball, and similarly *clearteammate(i)* is the clearance from the robot to the position of a teammate. Each robot should know its teammates score and defence costs. Each agent broadcasts its score and defence costs to its teammates. [8]

IV. EXPERIMENT

A. Approach

We will apply dynamic role assignment in Webots simulator. Goalkeeper is statically assigned to goal because of the rules of SPL. Goalkeeper uses a script that allow keeper only to stay put, wait and act only in case that ball is closer than 2 meters. He is able to walk sideways in direction of ball with limit of 8 steps where corner of the goal is. Is the ball reaches closer than 0.8m and angle between robot and ball is smaller than 10 degrees he will try to kick it out. We will try to assign two remaining field players as following. One of the players will be attacker and second defender. For estimation of costs we will consider distance and angle between robot and ball. We tried to include position of goal to final estimation but this caused

system overload and crashes. For costs estimation and communication is a must for bids to compare. After estimating the price both players will exchange their costs. Player with lower costs will be assigned as attacker and takes ball second one will be defender and will stay on home team half of field. We search for ball with simple algorithm. Robot sets his camera to upper position then turns round his head to capture pictures from horizon then he does the same for lower position of camera. If he fails to find ball he makes move backwards and repeats first move. If result is still negative, it will turn his body around 180 degrees and repeat both camera position scanning. If he fails again it will begin to do same moves from beginning. After he is successful in searching for ball distance and angle between direct position of player and ball will be estimated. We will estimate total cost as distance plus penalization for heading wrong direction. If angle is greater then 40 degrees we will penalize robot with coefficient 1, in case the angle is greater than 60 degrees we will use coefficient 1.3. These values were estimated as the same time it takes robot to turn as it takes to walk certain distance. For example if ball distance from robot is 1.456 m and angle between then is 68 degrees final costs will be 2.756. After exchanging total costs between players, robot with lower one will be attacker. It will approach ball depending on distance and angle. If distance is greater than 0.6m player will use 0.5m move forward. After that it slows down and if distance is smaller than 0.3m, it will switch in state of ball approaching. In this state robots goal is to reach position where angle between his direct position and ball is smaller than 10 degrees and angle between goal and his direct position is less than 20 degrees. Sometimes this inflicts very long and unwelcome “dances” around the ball.

B. Results

As shown on figure.1 first player (market with small Red Cross) will have to move backward to see the ball (red indicator). After this step he will be able to see the ball and has to wait for player two. In case of player two, simple move backwards, won’t be sufficient, After one step back it still won’t be able to see the ball, so it turns around 180 degrees (marked with blue arrow) to catch picture of ball. Figure.3 represents same situation after ball recognition and costs estimation.

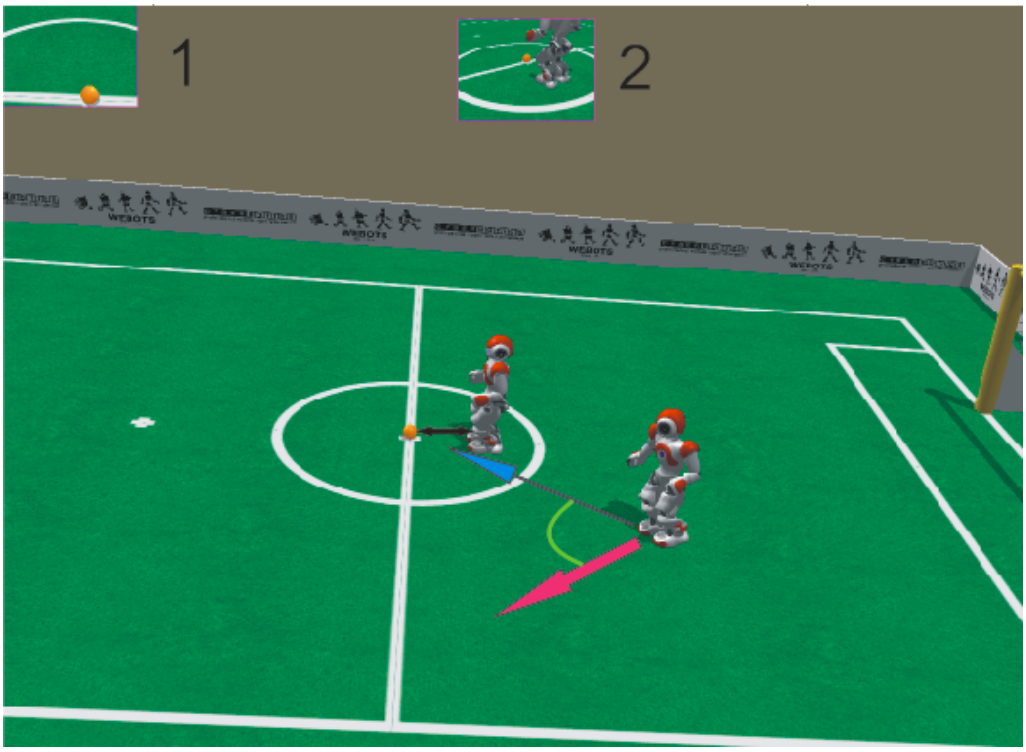


Fig. 3 After costs estimation

Small pictures in the top of figure.3 are from cameras of first player (marked 1) and second player (marked 2). 180 degrees move wasn’t very effective in this case because player two turn finished in angle bigger than 60 degrees from ball. The main disadvantage of simulator is that we can not stop Nao in middle of move because we risk that he will fall. So in this case we have lost about 4-5 second waiting for him to finish the move. In addition he is facing wrong

way so he will be penalized with coefficient 1.3 since angle between his direct position and ball is more than 60 degrees (green angle between red and blue arrow). Also player two should be penalized somehow for obstacle (player 1) in his way. As a judging factor for our experiments we take time. Some of the result under different conditions is show in table.1, in case that whole estimation took longer than 20 second we consider it as ultimate fail.

RE in table.1 stands for right estimation and RRE for right estimation in real world. Though algorithm was able to be successful in 3 from 5 tries for RE, for RRE was successful only in 2 cases and times were extremely high. RRE problem is shown in figure.4.

Table 1
Results

Nr.	Supervisor	RE	RRE	time
1	x	YES	YES	15.5
2	YES	no	no	x
3	x	no	no	>20
4	x	YES	YES	2
5	x	YES	no	1.3



Fig. 4 RRE

Even if player marked as 1 is closer to ball than player marked than 2 and algorithm will decide that player 1 should be attacker, which is correct by our approach. However, despite the fact that player marked as 2 is further than the player 1, the player 2 will be able to shot ball first into blue goalpost, because he has not to spin around the ball.

This problem and time results from other experiments show that this method suffers from lack of information. Since this and similar methods are applied in different leagues, where additional overhead camera is allowed. Even though this approach is successful in “lower” leagues it is obvious that is almost useless in SPL in Webots.

V. Conclusion

In this paper we researched methods currently used in strategy in different robotic soccer leagues based on MAS. We tested one of the methods as written above, under SPL standard in Webots simulator. Tested method has proven that is not very effective in this league because of luck of information that can be obtained thru Naos cameras. The main difference between ‘lower’ leagues and SPL is in overhead camera catching pictures from field and external computer to process images. Missing external computer turn not to be crucial for proposed method but it highly suffers from missing camera. We recommend focusing on more basic elements in future strategy development such as image processing, walking and upright

position control optimization, technique of shot and pass and location, rather then focus on higher skills as strategy or formations. Even latest games in [9] support this conclusion that in SPL league still more emphasis is to be on lower form of control.

ACKNOWLEDGMENT

This work is the result of the project implementation: Development of the Center of Information and Communication Technologies for Knowledge Systems (ITMS project code: 26220120030) supported by the Research & Development Operational Program funded by the ERDF.

REFERENCES

[1] Vidal, M. J. Fundamentals of Multiagent Systems
[2] Dorner ,J. Multiagentové systémy zamerané na riadenie mobilných robotických systémov
[3] Wooldridge, M. – Jennings, N. R.: Agent Theories, Architectures,and Languages: a Survey. Intelligent Agents. Ed.: Wooldridge, M.,Jennings, N. R. Berlin: Springer-Verlag, 1995
[4] http://www.robocup.org/robocup-soccer
[5] http://www.cyberbotics.com/products/webots/index.html
[6] Bruce, J. Zickler, S. Licitra, Veloso, M. CMDragons:Dynamic Passing and Strategy on a Champion Robot soccer Team, 2007
[7] http://robocup.bowdoin.edu/public/NorthernBitesReport2009.pdf
[8] KOSE, H., K. Kaplan, C. Mericli and H. L. Akin, "Genetic Algorithms Based Market Driven Multi-Agent Collaboration in the Robot-Soccer Domain", FIRA Robot World Congress 2003, October 1 - 3, 2003, Vienna
[9] http://www.cyberbotics.com/contest

Návrh riadenia činnosti práčky pomocou fuzzy kognitívnych máp

Ján VAŠČÁK

Centrum informačných a komunikačných technológií, Katedra kybernetiky a umelej inteligencie, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

jan.vascak@tuke.sk

***Abstrakt*—Táto práca sa zaoberá návrhom modelu adaptívneho regulátora s využitím fuzzy kognitívnych máp a hebbovského učenia pre účely riadenia práčky. Najprv si popíšeme fuzzy kognitívne mapy a niektoré metódy ich učenia. Ústrednou časťou tohto príspevku je rozšírenie konvenčného fuzzy regulátora pre použitie fuzzy kognitívnych máp s cieľom aplikovania hebbovského učenia optimalizáciu fuzzy produkčných pravidiel. Návrh vychádza z apriórnych znalostí experta pre proces prania, ktorým boli navrhnuté jednak štruktúra mapy a jednak počiatočne hodnoty. Následne je popísaná integrácia tejto štruktúry do regulátora a proces učenia. Záverom sú zhrnuté výsledky testov a vykonané ich hodnotenie metód učenia z hľadiska ich vhodnosti pre túto aplikáciu. Napokon sú načrtnuté budúce možnosti modifikácií a vylepšení tohto prístupu.**

***Kľúčové slová*—Fuzzy kognitívne mapy, adaptácia, práčka, hebbovské učenie.**

I. ÚVOD

Dnešný svet je natoľko zložitý, že je ťažké akýmkoľvek spôsobom popísať ktorýkoľvek reálny problém. Zmeniť sa to, pomocou rôznych metód, snažia mnohé tímy výskumníkov z celého sveta. Jednou z technológií, ktorá je značným prínosom v tejto problematike je aplikácia fuzzy logiky do konvenčných systémov. Hlavnou prednosťou fuzzy logiky je jej schopnosť pracovať s neurčitosťou, často používanou v ľudskom vyjadrovaní. Fuzzy logika umožňuje počítačom pochopiť výrazy ako „malý“, stredne starý“ apod. Prednosťou tejto technológie je aj vyhľadávanie informácií na základe nepresných alebo neúplných údajov aj s možnosťou nájsť chybne uložené informácie. Táto technológia má však aj svoje nevýhody a tou hlavnou je najmä neschopnosť učiť sa, t.j. vytvoriť si na základe fragmentov znalostí o skúmanom probléme celú bázu znalostí a neustále ju na základe vonkajších vplyvov externého sveta optimalizovať. Tento nedostatok je však možné odstrániť použitím tzv. hybridných systémov, ktoré sú už schopné samostatného učenia. Jednou z takýchto foriem je aj použitie fuzzy kognitívnych máp, ako nadstavby ku konvenčným fuzzy systémom.

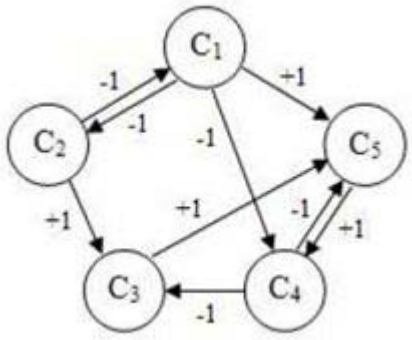
Pojem „Kognitívna mapa“ ako prvý použil americký psychológ Edward C. Tolman v roku 1948 v práci Cognitive Maps in Rats and Men [5]. V nej Tolman uvádza, že pri vnímaní okolitého prostredia dochádza nielen k okamžitému reaktívnemu chovaniu, ale v mozgu pokusných zvierat a zovšeobecnene aj v mozgu človeka sa vytvára virtuálny obraz vnímaného. Táto vnútorná prezentácia sa podľa Tolmana postupne upresňuje v priebehu učenia a tým umožňuje človeku (zvieraťu) orientovať sa v prostredí. Tolmanova teória kognitívnych máp vytrvala až do súčasnosti a patrí k základným východiskám kognitívnej psychológie.

Oživenie v oblasti kognitívnych máp priniesol v roku 1985 Bart Kosko v článku Fuzzy cognitive maps [1], v ktorom ukázal možnosti fuzzy rozšírenia pôvodných Axlerodových kognitívnych máp. Rozšírenia spočíva v tom, že kauzálne väzby medzi jednotlivými pojmami môžu nadobúdať hodnoty z intervalu <1;0>, resp. <-1;1> (viď obrázok 1).

Žiadna kognitívny mapa nie je systém, ktorý sa v čase nemení. Každý nasledujúci stav je vypočítaný z aktuálnych aktivačných hodnôt pojmov a spojení medzi nimi [9]. Dynamika správania sa fuzzy kognitívnych máp pozná 3 prípady:

- Prvým je prípad, kedy sa po konečnom počte prechodov stavový vektor už nemení. Je tiež označovaný ako pevný bod (fixed point – FP) podľa nemennej polohy stavového vektoru v stavovom priestore.

- Druhý prípad je označovaný ako medzný cyklus (limit cycle – LC). Nastáva v prípade, kedy po určitom počte prechodov dôjde k opakovaniu konečnej postupnosti stavov.
- Tretí prípad môže nastať pri komplexných FKM. Pre tento prípad sú charakteristické neperiodické zmeny stavového vektora, ktorý v stavovom priestore opisuje trajektóriu chaotického atraktora – CA [6].



Obr. 1 Fuzzy kognitívna mapa

Od doby vzniku pojmu FKM vzniklo veľa prác, ktoré si dali za cieľ popísať štruktúru a dynamické správanie FKM. Postupne sa vykryštalizovali dva základné prístupy:

- FKM ako rekurentná umelá neurónová sieť, prístup vychádzajúci z podobnosti medzi FKM a umelými rekurentnými neurónovými sieťami,
- FKM ako hybridný systém pravidiel a fuzzy množín, zdôrazňujúci pravidlový charakter väzieb v FKM.

V článku sa zameriam na druhú možnosť využitia FKM ako hybridného systému pravidiel a fuzzy množín.

II. VYBRANÉ METÓDY A PRÍSTUPY

Niekedy môže byť problémom zabezpečiť experta na vytvorenie komplexnej FKM z dôvodu vysokej zložitosti úlohy, ktorá presahuje ľudské možnosti. Najťažším krokom definovania FKM je učiť váhy medzi jednotlivými pojmami. Ak má FKM m vstupných a n výstupných pojmov, počet väzieb, ktoré je potrebné definovať je m*n. V prípade neschopnosti experta definovať spojenia z dôvodu veľkého počtu väzieb, je riešením použitie adaptívnych metód [9].

Existuje viacero metód adaptácie FKM, za všetky spomeniem tri, ktoré som si vybral pre adaptáciu riadiaceho systému práčky:

- Nelineárne Hebbove učenie
- Zdokonalené nelineárne Hebbove učenie
- Aktívne Hebbove učenie.

A. Nelineárne Hebbove učenie (NHL)

Algoritmus NHL je podľa [4] založený na predpoklade, že všetky koncepty FKM sú aktivované synchrónne v každom iteračnom kroku a rovnako synchrónne menia aj svoje hodnoty. Počas tohto procesu sú aktualizované všetky váhy w_{ij} prepojených konceptov a modifikovaná váha $w_{ji}^{(k)}$ je odvodená pre iteračný krok k.

V iteračnom kroku k + 1 je vypočítaná hodnota $A_i^{(k+1)}$ konceptu C_i, táto hodnota sa mení vplyvom hodnôt prepojených konceptov s hodnotami A_j vzhľadom na koncept C_i prostredníctvom upravených váh $W_{ji}^{(k)}$ v iteračnom kroku k.

Vzhľadom na využitie všeobecného nelineárneho Hebbovho učenia pre NN, bol zavedený matematický formalizmus začlenením tohto učenia pre FKM. Tento algoritmus sa týka hodnôt konceptov a hodnôt váh v modeli FKM a má tvar:

$$W_{ji}^{(k)} = \gamma \cdot w_{ji}^{(k-1)} + \eta A_i^{(k-1)}(A_j^{(k-1)} - \mathrm{sgn}(w_{ji}^{(k-1)}) w_{ji}^{(k-1)} A_i^{(k-1)}),$$

kde koeficient **η** je parameter učenia, stanovený experimentálnou metódou pokusu a omylu s cieľom optimalizovať výsledné riešenie a γ je parameter útlmu.

Implementácia NHL algoritmu vyžaduje určenie hornej a dolnej hranice učiaceho parametra η; pomocou experimentálnej metódy pokusu a omylu bol stanovený rozsah hodnôt učiaceho parametra η (0 ; 0,1).

V priebehu vývoja FKM odborníci definovali tzv. požadované výstupné koncepty (DOC). Tieto koncepty sú založené na základných charakteristikách a výstupoch systému, kde chceme odhadnúť ich hodnoty, ktoré odrážajú celkový stav systému. Rozlíšenie konceptov na vstupné a výstupné je stanovované expertmi pre každý špecifický problém.

Pre algoritmus NHL boli navrhnuté 2 kriteriálne funkcie:

1. $F_1 = \sqrt{\sum_{j=1}^{m}(DOC_j - T_j)^2}$,

Kde T_j je stredná cieľová hodnota výstupných konceptov DOC_j.

2. Druhá z kriteriálnych funkcií je daná odchýlkou po sebe nasledujúcich hodnôt DOC_j konceptov pri iteračnom kroku k podľa vzťahu:

$$F_2 = \left| DOC_j^{(k+1)} - DOC_j^{(k)} \right| < 0.002,$$

Kde DOC_j je hodnota j-tého DOC konceptu.

B. Zdokonalené Nelineárne Hebbove učenie (iNHL)

Ako aj v prípade algoritmu NHL aj v iNHL sú podľa [8] všetky koncepty aktivované synchrónne v každom iteračnom kroku a rovnako synchrónne menia aj svoje hodnoty. Sú aktualizované všetky váhy w_{ij} prepojených konceptov a modifikovaná váha $w_{ji}^{(k)}$ je odvodená pre iteračný krok k.

V iteračnom kroku k + 1 je vypočítaná hodnota $A_i^{(k+1)}$, táto hodnota sa mení vplyvom hodnôt prepojených konceptov s hodnotami A_j vzhľadom na koncept C_i prostredníctvom upravených váh $W_{ji}^{(k)}$ v iteračnom kroku k.

Na rozdiel od algoritmu NHL je tu zavedený tzv. impulzný parameter a navrhnuté pravidlo má tvar:

$$\Delta w_{ji}^{k} = \alpha_k \Delta w_{ji}^{k-1} + \eta_k z_k^2 (1 - z_k)(A_i^k - w_{ji}^{k-1} A_j^k),$$

kde $z_k = 1/(1+e^{-Aj(k)})$, koeficient α_k je impulzný parameter, ktorý má podobnú funkciu ako algoritmus back propagation u ANN a koeficient η_k, podobne ako pri algoritme NHL, je parameter učenia.

Učenie váh má tvar:

$$W_{ji}^{k+1} = w_{ji}^{k} + \Delta w_{ji}^{k}.$$

Opäť sú použité 2 kriteriálne funkcie:

1. Maximalizácia J, kde J tvoria hodnoty aktivácie výstupných konceptov (OC) a má tvar:

$$J = \sum_{i-1}^{N} (OC_i)^2,$$

kde N je číslo výstupného konceptu.

2. Druhou kriteriálnou funkciou je minimum odchýlky dvoch po sebe nasledujúcich hodnôt výstupných konceptov a má tvar:

$$\left| OC_j^{k+1} - OC_j^{k} \right| < e,$$

Kde **e** je maximálna povolená odchýlka, štandardne má hodnotu 0,001.

C. Aktívne Hebbove učenie (AHL)

Základným prvkom algoritmu AHL podľa [3] je stanovenie poradia aktivácie konceptov. V každom simulačnom kroku sa jeden alebo viacero konceptov stáva aktivačným konceptom, ktorý ovplyvňuje ostatné prepojené koncepty, ktoré sa v ďalšom iteračnom kroku sami stávajú aktivačnými konceptmi. Ak sa všetky koncepty stanú aktivačnými v zmysle stanoveného poradia aktivácie, aktivačný cyklus sa ukončí a začne sa nový aktivačný cyklus.

Okrem stanovenia poradia aktivácie konceptov je expertmi stanovený limitovaný počet konceptov ako výstupné pre každý konkrétny problém. Tieto koncepty sú definované ako aktivačné rozhodovacie koncepty (ADC). Tieto koncepty sú stredobodom nášho záujmu, pretože sú založené na základných faktoroch a vlastnostiach systému, známych ako výstupy, ktorých hodnoty odhadujeme a ktoré predstavujú konečný stav systému [3].

S cieľom zabrániť nekonečnému rastu hodnôt váh, odporúča sa podľa [4] normalizovať váhu na hodnotu 1, ||W|| = 1, v každom kroku aktualizácie:

$$W_{ji}(k) = \frac{(1-\gamma)w_{ji}(k-1) + \eta\, A_j^{act}(k-1).\, A_j(k-1)}{\left[\sum_{\substack{j=1 \\ j\neq i}} ((1-\gamma)w_{ji}(k-1) + \eta\, A_j^{act}(k-1).A_j(k-1))^2\right]^{1/2}}$$

Kde dodatok v menovateli pokrýva všetky prepojenia od aktivačných konceptov C_j k aktivovaným konceptom C_i.

Parametre učenia γ a η sú kladné hodnoty, ktorých veľkosť závisí na ohraničeniach a charakteristikách daného problému a ich hodnoty sa menia so zvyšujúcim sa počtom aktivačných cyklov podľa rovníc:

$\eta\,(c) = 0{,}02 \,.\, \exp(-0{,}1 \,.\, c)$

$\gamma\,(c) = 0{,}04 \,.\, \exp(-c)$,

kde c je index simulačného cyklu.

V algoritme AHL sú použité dve kriteriálne funkcie:

1. **Funkcia J**, ktorá sa zaoberá požadovanými hodnotami výstupných konceptov, ktorých hodnoty aktivácie konceptov nás zaujímajú. Má tvar:

$$J = ||ADC_i - A_i^{min}||^2 + ||ADC_i - A_i^{max}||^2 ,$$

kde A_j^{min} je minimálny cieľ hodnoty konceptu ADC_j a A_j^{max} zodpovedá maximálnej cieľovej hodnote ADC_j.

2. Odchýlka dvoch po sebe idúcich hodnôt daná vzťahom:

$$|ADC_j^{(c+1)} - ADC_j^{(c)}| < e = 0.001,$$

kde ADC_j je hodnota j-teho konceptu.

Cieľom tejto funkcie je ukončiť simulačný cyklus, ak je absolútna hodnota odchýlky ADC menšia ako e = 0,001. Ak sú počas simulácie obe kriteriálne funkcie splnené, je dosiahnutá konvergencia do rovnovážneho stavu a simulačný cyklus je ukončený.

AHL a NHL algoritmy nekontrolovaného učenia boli navrhnuté na jemné ladenie príčinných spojení FKM stanovených expertmi. Podľa [7] sú sprevádzané dobrou znalosťou daného systému alebo procesu a môžu prispieť k zavedeniu FKM ako robustnej metódy.

V nasledujúcej tabuľke uvediem základné rozdiely medzi Algoritmami AHL a NHL.

Tab. 1
Základné charakteristiky algoritmov AHL a NHL [7]

	AHL	NHL
1	Asynchrónna postupnosť aktivácie konceptov	Súčasné spustenie všetkých konceptov
2	Asynchrónna aktualizácia váh	Súčasná aktualizácia váh
3	Nové hodnoty pre všetky váhy	Aktualizované sú iba nenulové váhy
4	Vznik nových vzťahov medzi konceptmi FKM	Žiadne nové vzťahy medzi konceptmi FKM
5	Zavedenie aktivačného kroku a cyklu	Žiadne kroky, iba aktivačný cyklus
6	Dve kriteriálne funkcie založené na obmedzeniach konceptov	Dve kriteriálne funkcie pre algoritmus
7	Exponenciálny útlm parametrov γ a η	Konštantné hodnoty parametrov η a γ

V nasledujúcej kapitole popíšem riadiaci systém práčky pre ktorý som sa rozhodol implementovať fuzzy kognitívne mapy a zvolené algoritmy nekontrolovaného učenia AHL a NHL.

III. NÁVRH A IMPLEMENTÁCIA

Hlavným prínosom z aplikácie FKM v riadiacom systéme by mala byť optimalizácia výsledku pracieho procesu a tým aj úspora energie a vody. Aplikáciou metód učenia môže použitý regulátor v priebehu času optimalizovať parametre prania presne podľa vonkajších podmienok každého zákazníka. Táto unikátna vlastnosť využitím daných metód zároveň umožní ušetriť aj vývojové a laboratórne kapacity výrobcov práčok na určovanie optimálnych parametrov prania, ktoré sa pre konvenčný automat musia stanoviť pre široké spektrum

vyskytujúcich sa vonkajších faktorov vplývajúcich na výsledok pracieho cyklu, čo zvyčajne vyžaduje značné množstvo testov a skúšok.

Samotná aplikácia FKM a jej štruktúra závisí od vlastností riadeného systému a vlastností procesu prania. Tieto vlastnosti sú popísané v nasledujúcich kapitolách.

Pre stanovenie bázy pravidiel je dôležité popísať vlastnosti procesu prania, aby báza pokrývala všetky podstatné potreby pre zabezpečenie kvalitného výsledku.

Vlastnosti procesu prania

Pre dnešné automatické práčky je typické delenie pracieho cyklu, ktorý sa skladá z nasledujúcich činností:

- Predpranie
- Pranie
  - Zmáčanie
  - Fázu aktivácie enzýmov
  - Hygienickú fázu
  - Ochladzovanie
  - Vyčerpávanie
- Plákanie
- Odstreďovanie
- Sušenie

Každá časť cyklu má svoje zákonitosti a veľký priestor na optimalizáciu svojho priebehu v závislosti na type vloženej bielizne, na skutočnom znečistení a jeho množstve.

Pračka je zložená z týchto základných častí:

- Prací agregát práčky, zložený z nádrže a bubna poháňaného motorom. Svojou konštrukciou určuje základné charakteristiky riadenej sústavy. Obsahuje aj štandardný riadiaci modul motora práčky, ktorý poháňa bubon.
- Napúšťací systém, zabezpečujúci potrebný prívod vody a jej distribúciu v hydraulickom systéme práčky, obsahuje elektrické ventily na jeho ovládanie.
- Vypúšťací systém, vykonávajúci recirkulačný obeh a vypúšťanie vody z práčky, obsahuje elektrické čerpadlá, ktoré zabezpečujú prietok vody.
- Skriňa práčky, poskytuje priestor pre umiestnenie ovládacích prvkov práčky a potrebných vstupných a výstupných prvkov (vkladacie dvierka, násypka, prívodové hadice, výtokové hadice a podobne).

Už zo samotnej štruktúry riadenej sústavy je zrejmé, že jej presný matematický popis je veľmi zložitý, pretože v praxi ide o komplikovanú sústavu zloženú z mechanických, elektrických a hydraulických podsystémov, ktoré navyše pracujú vo veľmi širokom rozsahu prevádzkových parametrov. Toto by malo byť hlavným dôvodom na začatie používania adaptívnych systémov ako súčasti nadriadeného regulátora, zabezpečujúceho optimálne parametre pre prací cyklus, získané z priebežného učenia sa v reálnych podmienkach konkrétneho zákazníka.

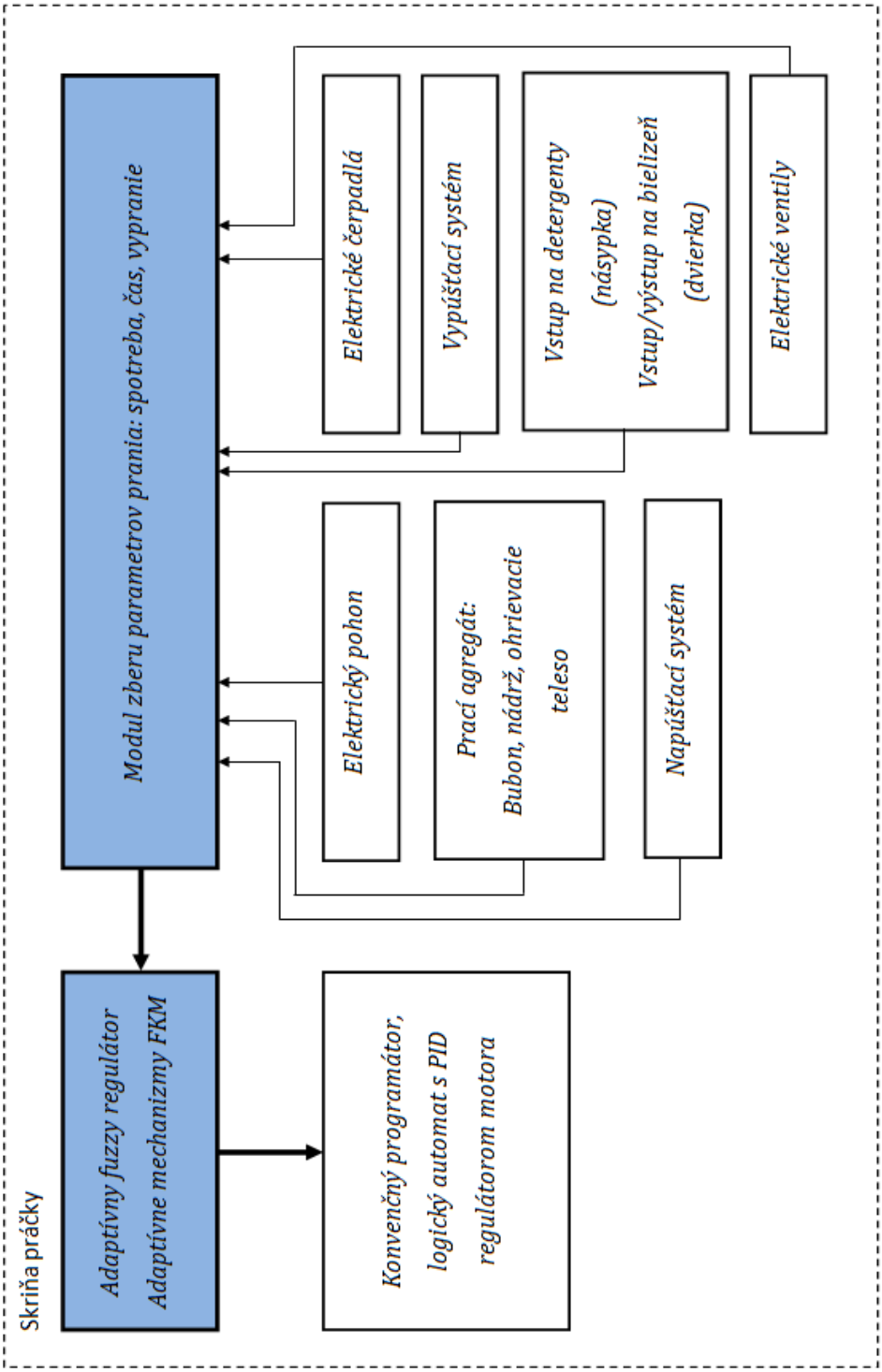
Aplikácia FKM v riadiacom systéme práčky

Cieľom navrhnutej štruktúry je optimalizovať činnosť konvenčného fuzzy regulátora ako nadstavby, ktorá svojimi vlastnosťami zlepší a optimalizuje parametre prania tak, aby zákazník v priebehu času získal optimálny výsledok prania pri čo najnižších nákladoch.

Celkovú štruktúru riadenia je preto možné popísať spôsobom, ako je to uvedené na obrázku 2, kde je znázornený zber dát s rôznych snímačov jednotlivých častí práčky do modulu pre zber dát, ktorý zozbierané dáta odosiela na spracovanie do adaptívneho fuzzy regulátora a ten vyhodnotené a spracované signály posiela programátoru, ktorý na základe týchto dát vykonáva samotné pranie.

Pre účely simulácie pracieho procesu bolo zvolené programovacie prostredie MS Visual Studio C#. Proces prania je ovládaný a optimalizovaný pomocou expertom navrhnutej FKM. Pre potreby použitia adaptívnych metód bolo potrebné získať množinu trénovacích dát. Trénovacie dáta pre adaptáciu riadiaceho systému práčky boli získané z bázy pravidiel stanovenej na základe expertných znalostí odborníkov z oblasti pracieho procesu. Na obrázku 3 je znázornená expertom navrhnutá FKM použitá na simuláciu procesu prania a jeho optimalizácie pomocou zvolených adaptívnych metód. Navrhnutá mapa obsahuje 22 výstupných pojmov, označených C_0 – C_{21}, a 9 vstupných pojmov, označených C_{22} – C_{30}.

Spojenia medzi jednotlivými pojmami boli expertmi stanovené na základe skúmania činnosti pracieho procesu v laboratórnych podmienkach pre dosiahnutie čo najväčšej efektívnosti použitia FKM.



Obr. 2 Štruktúra riadiaceho systému práčky

Vstupy do systému

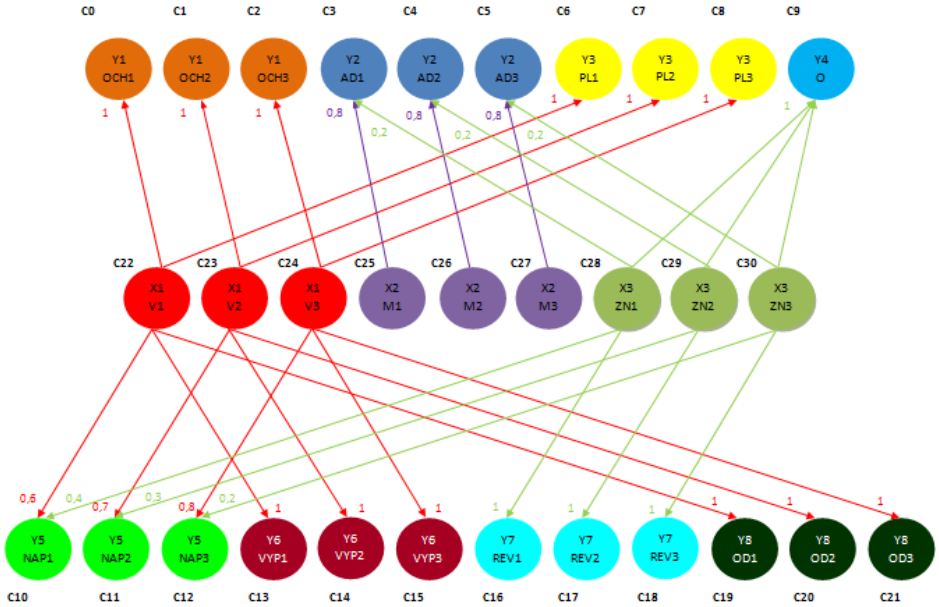
Boli zvolené tri vstupné parametre:

- Váha bielizne
- Stupeň mastnoty
- Stupeň znečistenia

Výstupné parametre

V závislosti od charakteristiky pracieho procesu boli zvolené nasledujúce výstupné parametre:

- Napúšťanie
- Ohrev
- Reverzácia
- Aktivácia detergentov
- Plákanie
- Ochladzovanie
- Vypúšťanie
- Odstreďovanie



Obr. 3 Expertom navrhnutá FKM

Jednotlivé hodnoty vstupných pojmov boli získané fuzzifikáciou hodnôt daných senzormi jednotlivých častí riadiaceho systému práčky a po vyhodnotení FKM sú defuzzifikované výstupné hodnoty použité na korekciu parametrov pracieho procesu.

Programová implementácia navrhnutého systému

Navrhnutý model riadiaceho systému práčky s využitím FKM a adaptačných metód NHL a AHL bol implementovaný a modifikovaný vo vývojovom prostredí MS Visual Studio v programovacom jazyku C# (obrázok 4). Vytvorená aplikácia vychádza z diplomovej práce „Adaptácia fuzzy kognitívnych máp pre potreby navigácie“ študenta Mareka Rutricha a ním vytvorenej aplikácie na ovládanie robota pomocou FKM [9].

Cieľom navrhnutej aplikácie bolo otestovať vhodnosť navrhnutých metód učenia FKM pre riadiaci systém práčky.

IV. Experimenty

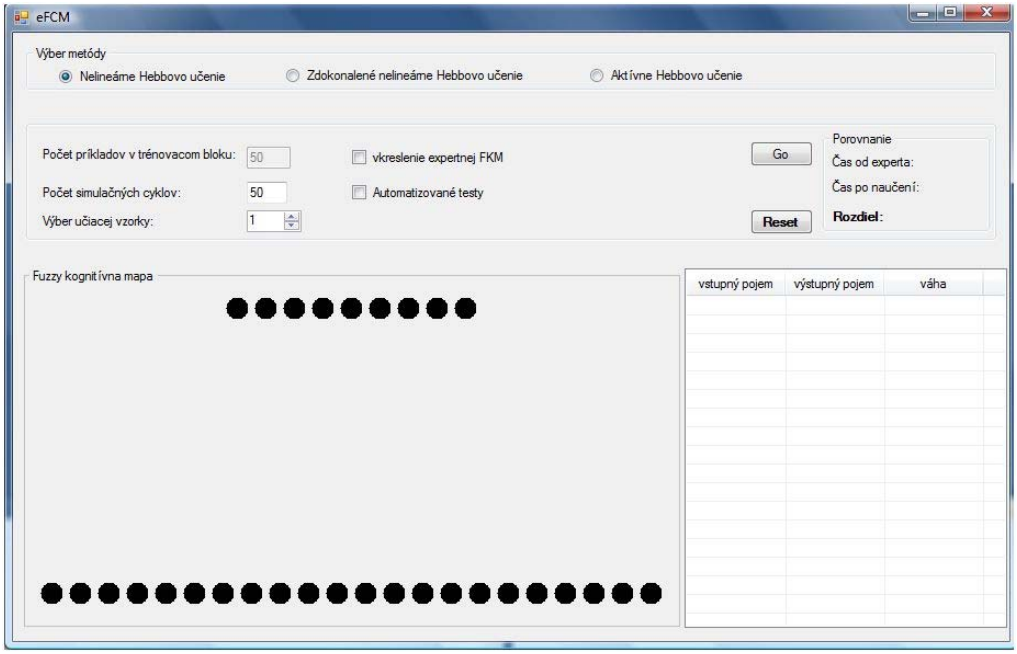
Pre účely tejto práce boli navrhnuté tri adaptívne úlohy pre návrh FKM, Nelineárne Hebbove učenie, zdokonalené nelineárne Hebbove učenie a Aktívne Hebbove učenie. Aby mohla byť vyhodnotená vhodnosť navrhnutých algoritmov, bolo potrebné vykonať na implementovaných algoritmoch testovanie. Vzhľadom na navrhnutú aplikáciu boli testy rozdelené na dve skupiny:

- **Testovanie na užívateľom zvolenej testovacej vzorke**

  Toto testovanie spočíva vo výbere jednej vzorky z celej množiny trénovacích dát a výbere počtu simulačných cyklov. Pre tento typ testov boli z celej množiny náhodne vybrané trénovacie vzorky pre každý z troch testovaných algoritmov.

  Pre testovanie boli náhodne z celej množiny 3000 trénovacích vzoriek vybrané trénovacie vzorky č. 1, 20, 100, 500, 1200, 2000 a 3000 pri počte simulačných cyklov 1, 15, 30 a 50.

- **Automatizované testy, kedy učenie prebieha na celej množine trénovacích dát**
  Táto forma testovania spočíva v učení na celej množine trénovacích dát s možnosťou výberu simulačných cyklov.
  Pri automatizovaných testoch bolo postupne vykonané učenie na 1000, 2000 a 3000 trénovacích vzorkách pri počte simulačných cyklov 1, 15, 30 a 50. Výsledky experimentov sú zobrazené v nasledujúcej tabuľke.



Obr. 4 Používateľské okno navrhnutej aplikácie

V. ZÁVER

Cieľom tejto práce bolo overiť vhodnosť navrhnutých adaptívnych metód učenia fuzzy kognitívnych máp, nájsť vhodnú metódu, schopnú nastaviť a optimalizovať parametre riadiaceho systému práčky, pre dosiahnutie optimálnych výsledkov pracieho procesu pri úspore vody a elektrickej energie.

Vhodnosť navrhnutých adaptívnych metód ukazujú experimenty vykonané pre jednotlivé metódy v kapitole 5. Výsledky testovania poukázali na rozdielnu účinnosť jednotlivých použitých metód, pričom ako najhoršia sa javí metóda nelineárneho Hebbovho učenia, ktorá vykázala najväčšie rozdiely v časoch výsledných dôb prania stanovených expertom a dôb prania po učení algoritmom NHL. Pri učení vznikli záporné väzby, ktoré ovplyvnili celkový výsledok negatívnym spôsobom. Metóda zdokonaleného nelineárneho učenia vykazovala už menšie, ale stále dosť rozdielne časy pracieho cyklu. Aj napriek skutočnosti, že pri tejto metóde sú aktualizované aj nenulové váhy konceptov, tieto vykazovali iba minimálne, zanedbateľné hodnoty, ktoré však aj napriek tomu mierne skreslili výsledné časy pracieho cyklu po učení. Najlepšie, aj keď nie ideálne sa javí metóda aktívneho Hebbovho učenia, kde jednotlivé časy pracieho cyklu sa najviac priblížili expertom stanoveným a ak vezmeme do úvahy podstatu týchto metód, ktorou je optimalizovať parametre bázy pravidiel, tak metódy do istej miery viac-menej túto skutočnosť naplnili.

Pre dosiahnutie ešte kvalitnejších výsledkov a dosiahnutia väčšej robustnosti riadiaceho systému práčky by som do budúcnosti navrhoval upraviť expertom stanovenú FKM, resp. rozdeliť jednotlivé fázy pracieho procesu na samostatné FKM, a nakoľko v tejto práci boli stanovené iba niektoré vstupné a výstupné parametre, ktoré nezohľadňujú celý rozsah parametrov, ovplyvňujúcich proces prania, rozšíriť bázu znalostí o kompletnú množinu parametrov vstupujúcich do procesu prania a parametrov z neho vystupujúcich.

POĎAKOVANIE

Táto práca bola vytvorená realizáciou projektu Rozvoj Centra informačných a komunikačných technológií pre znalostné systémy (kód ITMS projektu: 26220120030) na základe podpory

operačného programu Výskum a vývoj financovaného z Európskeho fondu regionálneho rozvoja.

LITERATÚRA

[1] KOSKO, Bart: Fuzzy Cognitive Maps. Int. Journal Man-Machine Studies [online], 1986 [cit. 2010-02-03], dostupné na internete: http://sipi.usc.edu/~kosko/FCM.pdf

[2] DICKERSON, Julie – KOSKO, Bart: Virtual Worlds as Fuzzy Cognitive Maps. The Massachusetts Institute of Technology [online], 1994, vol. 3, no. 2 [cit. 2010-02-03], dostupné na internete: http://sipi.usc.edu/~kosko/Virtual_Worlds_FCM.pdf

[3] PAPAGEORGIOU , Elpiniki – STYLIOS, Chrysostomos – GROUMPOS, Peter: Active Hebbian learning algorithm to train fuzzy cognitive maps. Int. Journal of Aproximate Reasoning [online], 2004 [cit. 2010-10-03], dostupné na internete: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.98.7416&rep=rep1&type=pdf

[4] PAPAGEORGIOU , Elpiniki – STYLIOS, Chrysostomos – GROUMPOS, Peter: Unsupervised learning techniques for fine-tuning fuzzy cognitive maps causal links. Int. Journal of Human-Computer Studies [online], 2006 [cit. 2010-02-03], dostupné na internete: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.95.9647&rep=rep1&type=pdf

[5] TOOLMANN, Edward C.: Cognitive maps in Rats and Men. In Classics in the History of Psychology. [online], 1948 [cit. 2010-02-03], dostupné na internete: http://psychclassics.asu.edu/Tolman/Maps/maps.htm

[6] MLS, Karel : Fuzzy kognitivní mapy (dizertačná práca), ČVUT Praha, 2006

[7] PAPAGEORGIOU , Elpiniki – STYLIOS, Chrysostomos – GROUMPOS, Peter: The Challenge of Using Unsupervised Learning Algorithms for Fuzzy Cognitive Maps : IEEE, 2004. s. 2425-2430

[8] LI, Sheng-Jun – SHEN, Rui-Min: Fuzzy Cognitive Map Learning Based on Improved Nonlinear Hebbian Rule : IEEE, 2004. s. 2301-2306

[9] Rutrich, Marek: Adaptácia fuzzy kognitívnych máp pre potreby navigácie. Diplomová práca. Košice: Technická univerzita v Košiciach, Fakulta elektrotechniky a informatiky, 2009. 71 s.

[10] Česánek, Štefan: Tvorba demonštračného programu činnosti fuzzy regulátora v automatickej pračke pre potreby výučby .Bakalárska práca. Košice: Technická univerzita v Košiciach ,Fakulta elektrotechniky a informatiky , 2008. 77 s.

Testovanie architektúr 1-Wire senzorových sietí

Jozef ADAMIŠIN, Ľudmila MACEKOVÁ

Katedra elektroniky a multimediálnych telekomunikácií, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

jozef.adamisin@student.tuke.sk, ludmila.macekova@tuke.sk

***Abstrakt* — Tento článok poskytuje opis princípov 1-Wire® senzorových sietí, so zvláštnym zameraním sa na ich topológie, a na ich hraničné možnosti. Je tu opísané testovanie niekoľkých návrhov topológií 1-Wire senzorovej siete v laboratórnych podmienkach, s využitím viacerých typov profesionálnych 1-Wire senzorov ako aj emulovaných teplotných senzorov riadených mikroprocesorom. Cieľom bolo navrhnúť stabilnú sieť s čo najväčším dosahom pri dostupnom výbere jej komponentov. Tieto experimenty, uskutočnené v rámci diplomových prác, umožnili praktické nahliadnutie na funkčnosť 1-Wire senzorovej siete, ktorá by mohla byť implementovaná do domácností, alebo automatizačných centier.**

***Kľúčové slová* — 1-Wire senzorová sieť, topológia, senzory**

I. ÚVOD

Užívatelia moderných budov požadujú čoraz väčší komfort a súbežne s týmito nárokmi rastú aj nároky na ekonómiu prevádzky takýchto objektov. Sú to nízkoenergetické budovy vybavené multifunkčnou technikou, a význam riadenia ich prevádzky sa zvyšuje s množstvom systémov, ktoré sa v objekte nachádzajú. V integrovanom systéme sú všetky podsystémy poprepájané a odovzdávajú získané informácie na vyhodnotenie centru, ktoré podľa optimalizovaného programu a aktuálnych potrieb obyvateľov rozhoduje o činnostiach jednotlivých podsystémov. Efektívnym riadením z centra možno zabezpečiť maximálnu účinnosť a minimálne náklady prevádzky takejto budovy. Jedným z riešení tejto problematiky je využitie 1-Wire [1] senzorovej siete.

1-Wire senzorová sieť je ideálna pre inštalácie, kde je požiadavka na použitie väčšieho počtu rôznych snímačov, aktuátorov a podobne, ktoré je potrebné koordinovať. 1-Wire technológia znamená aj využitie 1-Wire komponentov, ktoré sa vyznačujú nízkou cenou a jednoduchou inštaláciou. Táto technológia môže nájsť uplatnenie v priemysle, v domácnostiach, ale aj v rôznych zariadeniach spotrebnej elektroniky.

V laboratórnych podmienkach v rámci diplomových prác boli hľadané a testované hraničné možnosti architektúr 1-Wire sietí pri použití dostupných typov 1-Wire senzorov, káblov a ďalších komponentov.

V ďalších častiach tohto článku je uvedený stručný opis princípu, komponentov 1-Wire senzorovej siete, limity tejto siete, najčastejšie používané architektúry a typy napájania 1-Wire siete. Ďalej bude uvedený návrh experimentálnej 1-Wire senzorovej siete ako aj opis experimentov, ich výsledkov a záverov z nich vyplývajúcich.

II. PRINCÍPY A KOMPONENTY 1-WIRE SENZOROVEJ SIETE

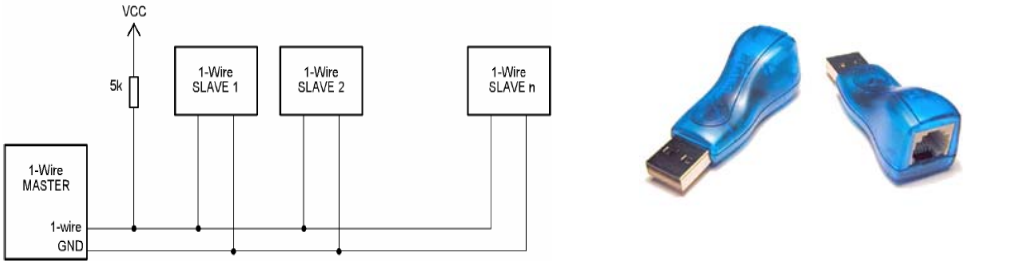
Technológia *1-Wire*® má pôvod v USA, v laboratóriách firmy Dallas Semiconductor, ktorá je od roku 2001 súčasťou firmy Maxim Integrated Products. 1-Wire sieť bola pôvodne bola vymyslená na komunikačné prepojenie zariadení na krátke vzdialenosti, typicky na vzdialenosti na plošnom spoji. Na komunikáciu používa iba jeden vodič. Komunikačný protokol je veľmi podobný protokolu I²C, no využíva menší dátový tok. V súčasnosti umožňuje prenos dát na väčšie vzdialenosti, a to rádovo až stovky metrov pri splnení určitých konštrukčno-realizačných podmienok.

Zbernica 1-Wire, ktorá môže mať niekoľko kusov až stovky odbočiek so senzormi, je v najjednoduchšej podobe zobrazená na Obr. 1. Vodič, spoločný pre komunikáciu aj pre napájanie celej siete je pripojený k riadiacemu zariadeniu (centru, označovanému *master),*

a zároveň k napájaciemu napätiu +5 V cez oddeľovací, tzv. pullup (zdvíhací) asi 5 kΩ rezistor v blízkosti mastra. Druhý vodič v tomto prípade slúži zároveň ako spoločný potenciál (zem, GND) aj ako návratový komunikačný vodič.

Riadenými zariadeniami, označovanými *slave,* sú senzory, spínače, príp. iné aktuátory a prevodníky s rôznymi funkciami. Špeciálnym typom slave zariadenia je tzv. *iButton*® [2], čo je pôvodný samostatný komponent od fy. Dallas s integrovaným obvodom v kovovom dotykovom púzdre TO92 veľkosti gombíkovej batérie, s funkciou odomykania, deaktivácie alarmu a podobne.

Zložitejšie siete, s dlhšími vodičmi, s väčším počtom slave zariadení, so slave zariadeniami vyžadujúcimi vlastné iné napájacie napätie, musia obsahovať aj napájací zdroj, čo znamená použitie viac než 2 vodičov.



Obr. 1. Základné zapojenie 1-Wire senzorovej siete s parazitným napájaním, a USB 1-Wire adaptér

1-Wire komunikácia je digitálna 2-úrovňová, unipolárna. Je založená na technológii zapojenia s otvoreným kolektorom, a na logike CMOS /TTL, čo pre vysokú úroveň znamená minimálne +2,2 V a pre nízku maximálne +0,8 V. Medzi týmito dvoma úrovňami je tzv. nedefinovaný stav, a prechod cez neho musí byť čo najrýchlejší. Logické nuly a jednotky ako aj signalizačné bity sú reprezentované nízkou (nulovou) úrovňou napätia trvajúcou rôzne dlhé, presne vymedzené časové úseky (*time slots* alebo *time chips*), po ktorých sa zbernica opäť čo najrýchlejšie vracia do nečinného (idle) stavu s „vysokou“ úrovňou napätia.

Dáta sú prenášané po metalických vedeniach, konkrétne po *krútených pároch*, ktorých konštrukcia (špirálovité krútenie s presným počtom krútení na jednotku dĺžky) oproti iným typom symetrických vedení výrazne potláča ovplyvniteľnosť vonkajšími rušeniami aj vyžarovanie do okolia. Ešte priaznivejší efekt odolnosti voči rušeniam majú káble s krútenými pármi so spoločným vonkajším tienením (FTP – káble) [7]. Usporiadanie vodičov UTP (Unshielded Twisted Pair) alebo FTP (Foiled Twisted Pair) káblov v konektoroch RJ45 resp. RJ12 je podľa štandardu T568A a ich určenie sa deje v zmysle odporúčaní pre 1-Wire [7].

1-Wire master ako jediné zariadenie môže inicializovať komunikáciu s 1-Wire slave zariadeniami. Úlohu riadiacej master jednotky môže plniť mikroprocesorový systém alebo osobný počítač, ktorý je k 1-Wire sieti pripojený cez tzv. *1-Wire adaptér*. Existujú adaptéry rôznej zložitosti pre sériové rozhrania RS-232, USB (Obr.1), aj pre paralelné rozhranie [2]. Adaptér musí poskytovať dostatočne veľký prúd na prekonanie celej dĺžky vedenia až k poslednému prídavnému zariadeniu. Taktiež musí generovať signál vhodného tvaru s časovaním podľa špecifikácie a optimalizovaný pre nabíjacie a vybíjacie časy v sieti. Posledná vlastnosť, ktorú musí adaptér spĺňať, je vhodné *impedančné prispôsobenie* (charakteristickej impedancii vedenia, čo je približne 100 Ω) tak, aby signály neboli odrazené späť do siete a nespôsobovali interferenciu s odozvou od iných 1-Wire zariadení.

Každé 1-Wire zariadenie má svoju *unikátnu adresu* danú vo výrobe 64-bitovým sériovým číslom. V jednom čase môže vysielať, resp. prijímať len jedno zariadenie a dáta sú prenášané po tom istom vodiči v obidvoch smeroch, nie však v tom istom čase (*half-duplex*). Komunikácia medzi dvoma slave zariadeniami nie je povolená. *1-Wire protokol* neumožňuje slave zariadeniu komunikovať, pokiaľ ho o to nepožiada riadiaca jednotka, čím sa predchádza kolíziám. Dátový prenos je realizovaný ako sekvencia bitov. Pripojené zariadenia majú vlastný hodinový generátor synchronizovaný riadiacou jednotkou. Zvláštny zdroj hodinových impulzov pre taktovanie zbernice nie je potrebný.

III. Architektúry 1-Wire senzorových sietí

Používajú sa nasledovné jednoduché typy architektúr, alebo ich kombinácie. [2]

Lineárna topológia: 1-Wire zbernica je tvorená jednoduchým párom vodičov. Začína v riadiacom zariadení a pokračuje do najvzdialenejšieho slave zariadenia. Ostatné slave

zariadenia sú pripojené k zbernici po jej dĺžke vetvami s minimálnou dĺžkou prípojného vedenia.

Vetvená topológia: zbernica je opäť, ako vyššie, tvorená jednoduchým dvojvodičovým vedením, ku ktorému sú ostatné slave zariadenia pripojené vetvami s dĺžkou 3 m a viac.

Hviezdicová topológia: 1-Wire zbernica sa rozdelí do viacerých vetiev v riadiacom zariadení, alebo v uzle čo najbližšie k riadiacemu zariadeniu. Vetvy hviezdy môžu mať rôznu dĺžku, a slave zariadenia môžu byť pripojené buď na konci vetiev, alebo pozdĺž vetiev v zmysle lineárnej topológie. Viac než pevná hviezdica sa však odporúča prepínanie jednotlivých vetiev hviezdice podľa potreby komunikácie vhodným spínačom, čo je spoľahlivejšie riešenie [2].

IV. Obmedzenia 1-Wire senzorovej siete

Pri návrhu 1-Wire sietí je potrebné brať ohľad na dva parametre siete, ktoré majú kritický vplyv na výkon 1-Wire siete. Týmito parametrami sú rádius a zaťaženie [2].

- *Rádius* je vzdialenosť najvzdialenejšieho slave zariadenia od master zariadenia, udávaná v metroch dĺžky metalického vedenia.
- *Zaťaženie* je súčet dĺžok všetkých dátových liniek udávaný v metroch.

Napríklad pri hviezdicovej topológii s troma vetvami dĺžok 10 m, 20 m a 30 m je rádius 1-Wire siete 30 m a zaťaženie je 60 m. Všeobecne platí, že zaťažiteľnosť siete je limitovaná nábehovým časom vo vedení (oneskorenie vplyvom vzdialenosti), pričom rádius určuje čas najpomalšej odozvy. Časové posuny odrazených signálov pri dlhej zbernici môžu spôsobiť zlyhanie komunikácie.

Druhým dôvodom zlyhania komunikácie býva degradácia logických úrovní signálu spôsobená paralelnými (aj parazitnými) kapacitami a sériovými rezistormi.

Každé slave zariadenie, prepojenie, spínač aj kábel pridáva do senzorovej siete záťaž, a zväčšuje tak jej celkovú časovú konštantu. Pri návrhu siete sa v dôsledku parazitnej kapacity dá prepočítať každé slave zariadenie na ekvivalentnú záťaž istého úseku vedenia, napr. približne 0,5 m alebo 1 m kábla. (napr. 50 pF zodpovedá približne 1 m vedenia typu FTP a pod.). Z toho potom vyplýva, že narastajúci počet použitých komponentov znižuje rádius siete (zo 750 m, v literatúre uvádzaných, na omnoho menej).

1-Wire adaptér môže výrazne ovplyvniť zaťažiteľnosť 1-Wire senzorovej siete, pripojenej k nemu [2]. Jednoduchý riadiaci adaptér obsahuje tranzistor (FET), ktorý vykonáva pulldown (sťahuje zbernicu do stavu log. 0), a pasívny rezistor, ktorý vykonáva pullup (zdvíha zbernicu do stavu log. 1). Jednoduchý pullup rezistor (Obr.1) má zaťaženie obmedzené asi na 200m kábla. Sofistikované 1-Wire master zariadenia používajú na prekonanie týchto limitov aktívne pullup obvody, ktoré poskytujú sieti vyššie prúdy a týmto spôsobom rozširujú maximálnu zaťažiteľnosť siete na viac ako 500m.

V. Spôsoby napájania 1-Wire senzorovej siete

Dátový vodič v 1-Wire senzorovej sieti štandardne dokáže zároveň napájať niekoľko menších senzorov, ako napríklad digitálne teplomery DS18S20 [4] a DS18B20 [5] a krátku sieť do niekoľkých metrov. Takéto napájanie sa označuje ako *parazitné.* Senzory obsahujú interné kapacitory (rádovo desiatky až stovky pF), ktoré sa v čase kľudu nabijú z dátového vodiča, a takto získanú energiu potom využijú v aktívnom stave.

Pre senzorovú sieť s väčším rádiom a zaťažením, alebo pri niektorých typoch slave zariadení, je nutné použiť prídavné stabilné napájacie napätie, prípadne aj viac typov napájacích napätí, ktoré sú zabezpečované z externého napájacieho zdroja.

VI. Návrh a realizácia 1-Wire senzorovej siete pre experimenty

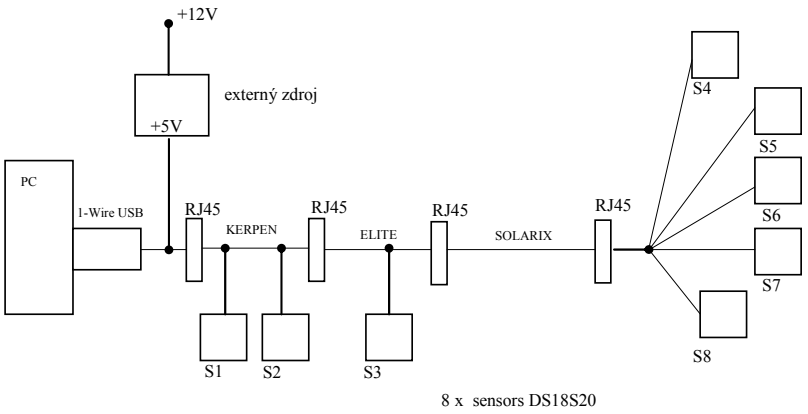
Cieľom našich laboratórnych experimentov bolo nájsť hraničné možnosti 1-Wire siete (dĺžky káblov niekoľkých typov, ktoré sme mali k dispozícii, počet a typy použitých senzorov zapojených s prispôsobovacími rezistormi alebo bez nich, rôzne typy prepojení), ako aj možnosti využitia a ďalšieho vylepšenia experimentálneho modulu s mikroprocesorom PIC16F628, emulujúceho 1-Wire teplotný senzor. Experimentálny emulačný modul bol navrhnutý v rámci inej diplomovej práce.

Na testovanie bolo navrhnutých niekoľko verzií architektúr siete. V podstate išlo o obmieňanie dĺžky lineárnej zbernice (56 až približne 370 m) s viacerými rozoberateľnými odbočeniami; prípadne sme na jej vzdialený koniec, alebo v inom, bližšom uzle, pripájali káble zapojené do hviezdy pozostávajúcej maximálne z 5 vetiev (Obr. 2).

Na prepojenie pri testovaní niekoľkých verzií architektúr siete boli podľa odporúčaní v [7] navrhnuté FTP káble kategórie cat5e, čo sú fóliou tienené káble pozostávajúce zo štvorice krútených párov. Bola použitá kombinácia rôzne dlhých úsekov FTP káblov značiek Kerpen, Elite

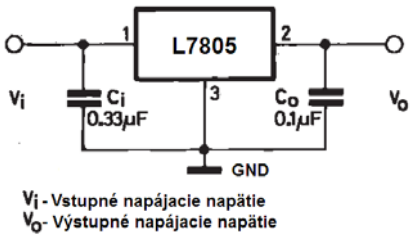
a Solarix, ktoré sme mali k dispozícii. Lacnejší telekomunikačný kábel s nekrútenými vodičmi sa pre tento účel ukázal ako nevhodný. 1-Wire sieť s jeho použitím pracovala len do dĺžky približne 20m.

Na realizáciu variabilného predlžovania, odbočovania zbernice, aj rozvetvovania do hviezdy, po skúmaní niekoľkých iných riešení, ktoré sa ukázali ako nespoľahlivé, nepraktické alebo drahé a v malých počtoch nedostupné (10 pinové IDC konektory, pevné spájkované spoje, alebo telekomunikačné spojovacie prvky typu „lentilka“ UR2 od americkej fy. 3M alebo podobné ABECOMSA od juhoafrickej fy. Dryden Comm., Ltd.) sme navrhli štandardizované konektory a spojovacie členy technológie RJ45 (RJ45-CMB combiner, resp. splitter RJ45 2x8/8F). Na realizáciu rozoberateľného pripojenia senzorov sme použili cca. 10 cm úseky plochého telekomunikačného kábla a konektory typu RJ45 zapojené podľa odporúčaní fy. Dallas [7]. Pre potreby snímania hodnôt z asi pätnástich bytových senzorov, ako sú teplomery, príp. vlhkomery a podobne, a pre externé napájanie emulačného modulu bolo pre 1-Wire senzorovú sieť navrhnuté a realizované využitie 3 vodičov z FTP-kábla, a to: 1-Wire dáta – modrý vodič, napájanie +5V DC – zelený vodič, GND – zeleno-biely vodič.



Obr. 2 Realizovaná kombinovaná 1-Wire architektúra – lineárna zbernica plus hviezda - s vyznačením použitých typov FTP káblov, senzorov a prepojovacích komponentov pre rozoberateľné prepojenia

Externý zdroj +5 V bol a realizovaný v podobe jednoduchého zapojenia so stabilizátorom L7805 (Obr.3). [6].



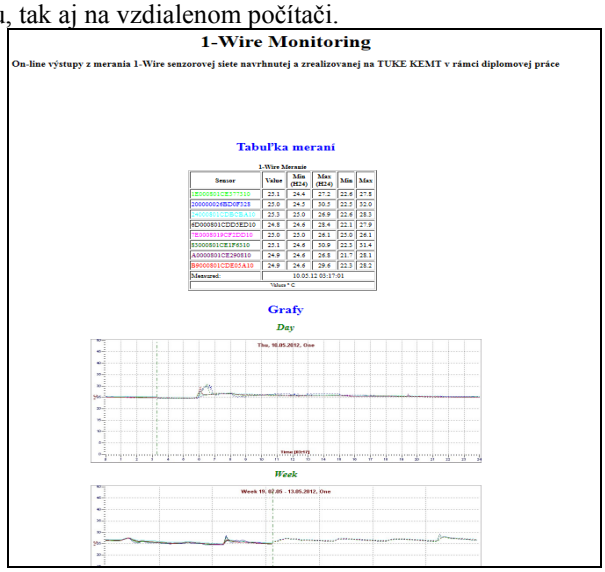
Obr. 3. Napájací obvod so stabilizátorom L7805 pre +5V

Do siete pri mnohých pokusoch boli nakoniec pripájané v rôznych kombináciách poradia a vzdialenosti väčšinou iba teplotné senzory typu DS18S20 ako aj niekoľko vývojových verzií modulu s mikroprocesorom PIC16F628 emulujúceho teplotný senzor s nastaviteľnými parametrami.

Nebol použitý žiadny mechanizmus riadenia nábežných a zostupných hrán impulzov komunikačného priebehu (slew rate control).

Ako master zariadenie pre všetky experimenty slúžil osobný počítač (desktop alebo laptop s verziou OS Windows XP, príp. vyššou) s adaptérom 1-Wire USB typu DS9490R [8] od firmy Maxim Integrated Products. Pre sledovanie a zber dát môže poslúžiť program OneWireViewer [9], alebo program LogTemp [10]. My sme sa rozhodli pre LogTemp; ním bolo možné sledovať viac senzorov súčasne, nastaviť vhodné grafické zobrazenie výsledkov (malá ukážka na Obr. 4), alarmy, príp. odosielanie e-mailových správ s chybovými hláseniami. Niekoľko týždňov kontinuálne prebiehajúce merania sa v pravidelných intervaloch zaznamenávali v podobe grafov, tabuľky aj číselných textových súborov. Uvedené výstupy bolo možné sledovať aj pomocou verejne prístupnej web-stránky. Niektoré parametre merania bolo možné podľa potreby

nastavovať na diaľku s využitím programu TeamViewer nainštalovaného ako na riadiacom počítači v laboratóriu, tak aj na vzdialenom počítači.



Obr. 4 Ukážka výstupov z programu LogTemp, sledovaných na web stránke: vizualizácia kontinuálnych meraní teplôt pomocou 1-Wire siete obsahujúcej 8 teplotných senzorov.

Analýza správania sa siete bola uskutočňovaná sledovaním jej funkčnosti a stability pomocou spomínaných softvérov, a v prípade problémov aj vyhodnocovaním časových a napäťových parametrov komunikačnej sekvencie pomocou osciloskopu. V prípade osciloskopických meraní bola sieť riadená z laptopu napájaného batériou, aby sme sa vyhli prípadným problémom s prúdmi tečúcimi nulovým vodičom, ktorý bol spoločný pre napájanie aj pre komunikáciu.

VII. VÝSLEDKY EXPERIMENTOV

Po mnohých rôznych obmenách zapojenia siete, senzorov a emulačných obvodov, pripojených k nej, môžeme sformulovať nasledujúce prakticky získané, resp. overené poznatky:

- Overili sme, že pri dlhej sieti (300 m a viac) celkový sériový odpor prekračujúci istú hodnotu môže sieť znefunkčniť. V našom prípade to spôsobil ochranný rezistor emulačného modulu, ale aj sériové prispôsobovacie 100-, resp. 120 ohmové rezistory zapojené vo vetvách so senzormi vo veľkých vzdialenostiach od začiatku siete (56 m a viac). Priebeh 1-Wire komunikácie sa odklonil od prípustného tvaru v podobe „nedotiahnutých“ núl alebo/aj jednotiek (Obr. 5).
- Dlhý kábel (veľký rádius siete), kábel s veľkou dĺžkovou kapacitou, ale aj zariadenie s dlhou dobou odozvy (dlhšou, než pripúšťa 1-Wire protokol [7]) môžu spôsobiť výpadok 1-Wire siete alebo jej nestabilitu.
- Ako jedna z možných príčin problémov sa v originálnej (Maxim Integrated Products) literatúre uvádzajú podkmity, doznievanie a prekmity komunikčného priebehu, spôsobené mnohonásobnými odrazmi na impedančne neprispôsobených odbočeniach vedenia. V tejto súvislosti, aj keď sa takéto skreslenia objavili, sme pri komunikácii nezaznamenali problémy, pokiaľ sa nevyskytli súčasne s veľkým oneskorením, alebo pomalým nábehom, resp. dobehom impulzov.

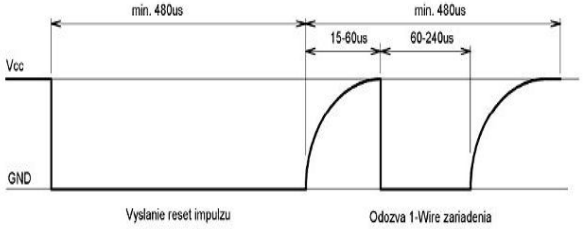
VIII. ZÁVER

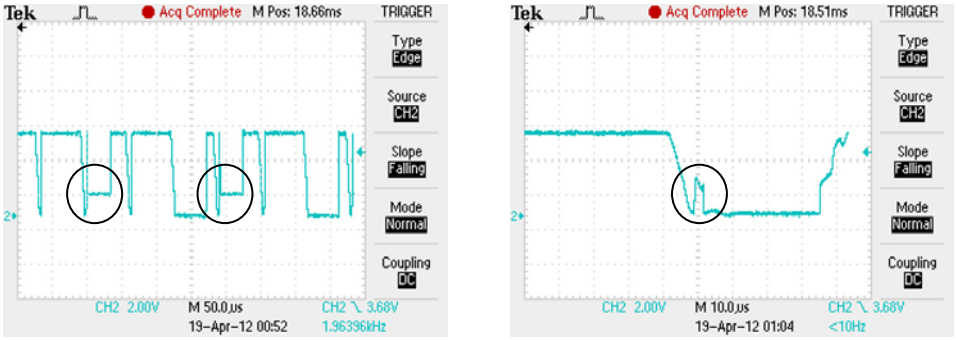
V príspevku sme opísali naše skúsenosti s experimentálnou realizáciou 1-Wire siete, približujúcou sa skutočnej realizácii, čo sa týka dĺžky káblov siete, množstva a typov senzorov, rozhrania a spôsobu riadenia. Hlavnou skúsenosťou možno nazvať poznatok, že tento typ siete je pomerne jednoducho realizovateľný vďaka štandardizovaným 1-Wire komponentom, a nenáročný aj z pohľadu ceny. Chybám, zlyhaniam, neistotám, a nespoľahlivosti sa dá predísť starostlivým teoretickým návrhom, ktorý je výhodnejší než množstvo neúspešných pokusov.

POĎAKOVANIE

Podporujeme výskumné aktivity na Slovensku/ Projekt je spolufinancovaný zo zdrojov EÚ. Tento článok bol vypracovaný v rámci projektu "Centrum excelentnosti integrovaného výskumu a využitia progre-

sívnych materiálov a technológií v oblasti automobilovej elektroniky", ITMS 26220120055.





Obr. 5 Pre porovnanie: ukážka predpísaných úrovní 1-Wire komunikačných bitov (hore) [4], a namerané priebehy pri nevhodne zaťaženej 1-Wire sieti (veľký sériový odpor).





Obr. 6. Fotografie pracoviska s laboratórnou 1-Wire sieťou (v krabici je 305m dlhý FTP Solarix kábel)

LITERATÚRA

[1] Maxim Integrated Products: 1-Wire Devices. Dostupné na http://www.maxim-ic.com/products/1-Wire
[2] Application Note AN148: Guidelines for Reliable Long Line 1-Wire Metwork. Maxim. Dostupné na http://pdfserv.maxim-ic.com/en/an/AN148.pdf
[3] Kábel 5. Kategórie, Wikipédia. Dostupné na http://en.wikipedia.org/wiki/Category 5 cable
[4] DS18S20 Programmable Resolution 1-Wire Digital Thermometer, Maxim Graphics. Dostupné na http://datasheets.maxim-ic.com/en/ds/DS18S20.pdf
[5] DS18B20Programmable Resolution 1-Wire Digital Thermometer, Maxim Graphics. Dostupné na http://datasheets.maxim-ic.com/en/ds/DS18B20.pdf
[6] L78XX, Positive voltage regulators. Dostupné na http://www.st.com/stonline/stappl/st/com/TECHNICAL_RESOURCES/TECHNICAL_LITERATURE/DATASHEET/CD00000444.pdf
[7] Springbok Digitronics: Innovative Hardware and Software for the Modern World – 1-Wire FAQ, Design Guide v1.0, 8/19/04.
[8] OneWireViewer User´s Guide, Version 1.5, Maxim. Dostupné na http://pdfserv.maxim-ic.com/en/an/AN3358.pdf
[9] DS9490R/DS9490B USB do 1-Wire/iButton Adapters, Maxim. Dostupné na http://pdfserv.maxim-ic.com/en/ds/DS9490-DS9490R.pdf
[10] LogTemp - program, © 2011 MR Soft Tmi. Dostupné na http://www.mrsoft.fi/ohj01en.htm

Základné princípy MIMO systému pre komunikáciu v automobilovej doprave

Dušan DEMIAN, Iveta GLADIŠOVÁ

Katedra elektroniky a multimediálnych telekomunikácií, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

dusan.demian@gmail.com, iveta.gladisova@tuke.sk

***Abstrakt* — Článok popisuje súhrn, štandardy a trendy bezdrôtových komunikačných technológií v oblasti automobilovej dopravy. Dôležitým faktorom pre súčasnosť, ale aj budúcnosť v tomto uponáhľanom svete je zvýšenie bezpečnosti a komfortnosti ľudí vo vozidle a na vozovke. Pre spôsob komunikácie v automobilovej doprave je možné využívať MIMO systém, ktorý je opísaný z hľadiska jeho základných princípov. Ďalej sa popisuje model MIMO kanála, jeho matica a kapacita. Aplikovaním MIMO systému dochádza k zvýšeniu dátovej priepustnosti, výkonu, dosahu vysielania, zároveň zvýšeniu počtu užívateľov bez rozšírenia šírky pásma na základe väčšej spektrálnej účinnosti. V automobilovej doprave sa pre komunikáciu využívajú aj technológie Wi-Fi (štandard 802.11p) a WiMAX (štandard 802.16e), ktoré sú stručne charakterizované.**

***Kľúčové slová* — V2V2I, MIMO technológia, diverzita, priestorovo-časové kódy, kapacita kanála**

I. Úvod

Komunikačné systémy v automobilovej doprave sú predmetom záujmu výskumných a vývojových skupín na celom svete už skoro desať rokov. Neustály rozvoj dopravy, zvyšovanie sa premávky, ako aj integrácie inteligentných komunikačných systémov do automobilov, či väčších komponentov so sebou prináša i problém jednoduchej a bezpečnej komunikácie vozidiel. V priebehu desiatich rokov boli analyzované služby, ktoré môžu byť úspešne realizované s využitím systému V2V (Vehicle-to-Vehicle), komunikačné spojenie medzi vozidlami navzájom a V2I (Vehicle-to-Infrastructure), komunikačné spojenie medzi vozidlami a infraštruktúrou (niekedy sa zvykne používať aj skratka V2V2I) [1], [2]. Boli vyvinuté vhodné komunikačné štandardy, vypracované návrhy na vyhradenie frekvenčných pásiem a realizované prvé prototypy komunikačných jednotiek [7], [8]. Jednou z mála sporných otázok je nedostatočná koordinácia medzi jednotlivými regiónmi, ktoré sa uvedenou problematikou intenzívne zaoberajú (napr. USA, Japonsko a Európa).

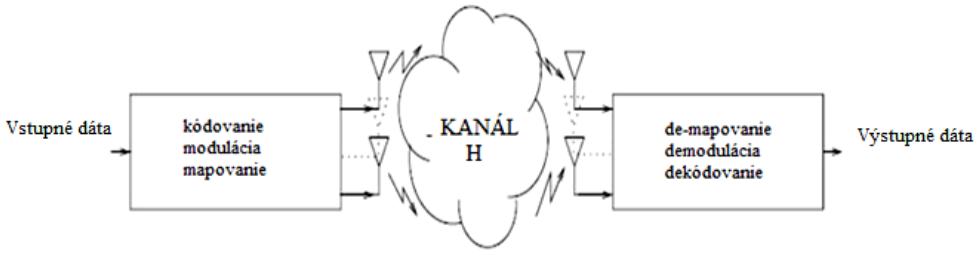
Napriek niektorým problémom s harmonizáciou využívaných a navrhovaných štandardov, s otvorenými otázkami zabezpečenia správ a dôvernosti informácií, je možné predpokladať, že v priebehu nasledujúcich rokov sa na európskych cestách budeme čoraz častejšie stretávať s vozidlami, ktoré sa budú vedieť medzi sebou dorozumieť a následne informovať vodiča o hroziacej nehode, obmedzenej rýchlosti, nízkej cene paliva, voľných parkovacích miestach, prípadne o prekážke na nasledujúcom úseku komunikácie, atď. [10]. V súčasnej dobe táto problematika patrí k dynamicky rastúcej oblasti. Oprávnene môžeme teda očakávať, že v najbližších rokoch komunikačné systémy v automobilovej doprave významnou mierou prispejú k zvýšeniu spoľahlivosti, bezpečnosti, komfortnosti a efektívnosti cestnej premávky.

II. MIMO technológia

Požiadavky a náročnosť v automobilovej bezdrôtovej komunikácii neustále narastá, čo má negatívny dopad na frekvenčné pásmo, útlm, viaccestné šírenie, prenosovú rýchlosť atď. Aby sa zabezpečil určitý štandard služieb, je dôležité riešiť tento problém. Ako vhodné riešenie sa javia systémy MIMO (Multiple-Input-Multiple Output), ktoré z hľadiska zvýšenia spektrálnej účinnosti môžu byť nástupcami dnešných komunikačných systémov.

MIMO technológia používa niekoľko vysielacích a prijímacích antén v multi-anténnom komunikačnom systéme (Obr. 1). Kódovanie, modulácia a mapovanie signálov v jednotlivých anténach môžu byť realizované spoločne alebo osobitne. Aplikovaním MIMO systému v bezdrôtovej komunikácií dochádza k zvýšeniu dátovej priepustnosti, dosahu vysielania bez

rozšírenia šírky pásma a vysielacieho výkonu. Dosahujú to pomocou väčšej spektrálnej účinnosti, vyššou spoľahlivosťou prenosu a znížením útlmu [3], [4].



Obr. 1 Bloková schéma bezdrôtového MIMO systému

Súčasné aj budúce bezdrôtové komunikačné MIMO platformy sa môžu vyznačovať rôznymi vlastnosťami a to napr.

- prostredím (interiér, exteriér),
- bezdrôtovým štandardom,
- nosnou frekvenciou,
- šírkou pásma ,
- vzorkovacou frekvenciou,
- počtom antén,
- spätnou väzbou alebo bez spätnej väzby,
- počtom užívateľov - jeden alebo viac,
- druhom modulácie,
- základnými algoritmami,
- dosiahnuteľnou prenosovou rýchlosťou,
- dosahom.

Zrejmé je, že s jednou platformou, ktorá má najvyššiu flexibilitu a modularitu nemusíme dosiahnuť pozitívne výsledky všetkých vlastností. Potrebné je zistiť vlastnosti ako flexibilita, škálovateľnosť a výkonnosť danej platformy [5]. Akonáhle je určené prostredie, tak poznáme štandard, rozsah nosných frekvencií, šírku pásma, čo zjednodušuje ďalší výber vlastností, ako napr. typ antén.

Čo sa týka platformy pre vonkajšie prostredie, tu je testovanie mobilnými zariadeniami ekonomicky a časovo náročnejšie a preto dostupnosť štandardov je obmedzená. Pri tomto type platformy jedným z kritérií pre „dokonalý“ prenos je vhodný výber anténnej sústavy. Vzdialenosť jednotlivých antén je menšia alebo rovná ako polovica vlnovej dĺžky t.j. $\lambda_c/2 = c_0/f_c$ (to je približne 1 cm pri 14 GHz a 7,5 cm pri 2 GHz), f_C je nosná frekvencia. Primerané hodnoty nosnej frekvencie sú v pásme ISM 2,4 GHz a 5,2 – 5,9 GHz. Tieto frekvencie zodpovedajú bezdrôtovým sieťam štandardu IEEE 802.11 alebo IEEE 802.16. Navrhnutá platforma musí podporovať oba prenosové smery, spätnú väzbu kanála pre kontrolu informácie, nastaviteľnú vzdialenosť antén a typ modulácie. Musí byť schopná spracovať príslušné triedy algoritmov, ako je priestorové multiplexovanie, tvarovanie lúča, priestorovo-časové kódovanie a akúkoľvek ich kombináciu [5]. Požiadavky pre návrh technických a metodických vlastností systému bezdrôtových MIMO prototypov sú náročné pre veľkú zložitosť spracovania algoritmov signálu. Zvyčajne sa rozdelia do niekoľkých algoritmov v DSP (Digital Signal-Processing) a FPGA (Field-Programmable-Gate Array)**.**

Čo sa týka technológii, tak technológia Wi-Fi na prenos dát vo V2V2I komunikácii [7] využíva spoluprácu základného štandardu (802.11) a novších verzií napr. 802.11p. Komunikáciu medzi vozidlami na kratšiu vzdialenosť zabezpečuje spojenie krátkeho dosahu tzv. DSRC, ktoré využíva frekvenčné pásmo v rozmedzí 5,85 až 5,925 GHz. Dáta sa môžu prenášať rýchlosťou až 27 Mb/s na šírke pásma 10MHz. Pre zachovanie spomenutých vlastností je veľmi dôležitý návrh anténneho systému. Pre tento štandard sa osvedčilo použitie modulov zložených s jednopólových a mikropásikových antén, ale vzhľadom na ich umiestnenie na vozidle. Mikropásikové (patch) antény je vhodné umiestniť na predných spätných zrkadlách, pretože dosahujú lepší príjem v prípade približujúceho sa vozidla. Jednopólová anténa umiestená na strecha má lepší príjem signálu, keď sa vozidlá míňajú resp. vzďaľujú. Anténne moduly najčastejšie vyžívajú kombináciu 2x2, 3x3 alebo 4x4 MIMO antén [6], [7].

Technológia WiMAX využíva pre prenos dát vo V2V2I komunikáciách základný štandard (802.16) v spolupráci s novšími variantmi napr. 802.16e. Dáta sa šíria v pomerne širokom frekvenčnom pásme 2 až 66 GHz v šírke pásma od 1,25 do 20 MHz. WiMAX technológia pre komunikáciu by mohla nahradiť vo vonkajšom prostredí technológiu Wi-Fi. Ideálnym prípadom by bola spolupráca týchto technológií pre zlepšenie prenosu dát. Tento systém umožňuje prenášať dáta teoretickou prenosovou rýchlosťou až 15Mb/s pri použití 5 MHz kanála. Pokrytie územia dosahuje v najlepších podmienkach až 50 km pri priamej viditeľnosti a

už len 8 km pri nepriamej viditeľnosti antén. Návrh anténnych systémov v tomto prípade ovplyvňujú charakteristiky, parametre a požiadavky [9].

III. PRINCÍPY MIMO SYSTÉMOV

Za účelom využitia možností, ktoré ponúka prítomnosť viacerých prijímacích a vysielacích antén v MIMO systémoch boli vyvinuté základné princípy: diverzita, časovo-priestorové kódy a modulácia

Diverzita vo všeobecnosti používa určitý počet oddelených prenosových kanálov z dôvodu zníženia až potlačenia únikov za účelom vylepšenia spoľahlivosti komunikačných liniek. Vysielaním viacerých kópii signálu cez rôzne komunikačné kanály sa výrazne zníži možnosť znehodnotenia signálu vplyvom únikov (v prípade šírenia signálu cez kanál s hlbokým únikom môže dôjsť k jeho strate či pohlteniu). Diverzita je získanie kópii rovnakej informácie na strane prijímačov. Hlavnou úlohou je teda vysielanie a prijímanie signálov cez viacero kanálov, aby nedošlo k ich strate. V MIMO systémoch sa využíva diverzita pre lepšie využitie kapacity prenosového bezdrôtového kanála medzi vysielačom a prijímačom. Táto diverzita je označovaná ako priestorová (signály sú nezávislé, nakoľko antény sú v určitej vzdialenosti, ktorá závisí od frekvencie, prostredia a výšky), podieľa sa na znížení únikov a zlepšení kvality kanála. Diverzitu môžeme rozdeliť podľa usporiadania antén na [8]: diverzita na strane prijímača, diverzita na strane vysielača a ich kombinácia.

Medzi techniky, ktoré vytvárajú diverzitu patrí časovo priestorové kódovanie. Medzi známe *priestorovo –časové kódy* patrí:

• MRRC (Maximal-Ratio-Receiver-Combining) – základná metóda získavania diverzity. Tradičný spôsob kombinovania prijatých signálov z rôznych antén využívaný aj v GSM (Global System for Mobile Communications) bázových staniciach.

• Alamoutiho schéma – predstavuje jednu z najjednoduchších schém, ktorú možno aplikovať na väčší počet vysielacích antén. Na bázovej stanici sú umiestnené dodatočné antény, ktoré využívajú všetky komunikačné terminály. Poskytuje plnú diverzitu a vyžaduje iba lineárne operácie vo vysielači a prijímači.

• STBC (Space-Time-Block-Code) – priestorovo-časové blokové kódy. Kódovanie pomocou matice kódu. Dekódovanie je výberom symbolov, ktoré boli vyslané s najväčšou pravdepodobnosťou porovnaním s lineárnou kombináciou prijatých signálov v priestore a čase.

• STTC (Space-Time-Trellis-Code) – priestorovo-časové mriežkové kódy. Mapovanie vstupu do prechodov medzi stavmi v prechodovom diagrame. Princíp je podobný konvolučnému kódu, avšak jeden stav je definovaný cez viacero vysielacích antén. Dekódovanie je možné aj výberom najpravdepodobnejšej možnosti alebo Viterbiho dekóderom .

Modulácia OFDM. Pri vývoji systémov MIMO - OFDM (Orthogonal-Frequency-Division-Multiplexing) sa prihliadalo na kompromis medzi parametrami: rýchlosťou, dosahom a spoľahlivosťou. Pretože rýchlosť by mohla byť zvýšená len na základe zníženia dosahu a spoľahlivosti, zatiaľ čo dosah by mohol byť zvýšený iba na úkor rýchlosti a spoľahlivosti. Spoľahlivosť by sa mohla zlepšiť tým, že sa zníži rýchlosť a dosah. Modulácia MIMO-OFDM zadefinovala tieto vzťahy a z nich vyplývajúce kompromisy, čím jasne potvrdila, že môže zlepšiť všetky parametre súčasne. Rieši problémy spôsobené frekvenčnou selektivitou kanála a medzisymbolovým rušením [3].

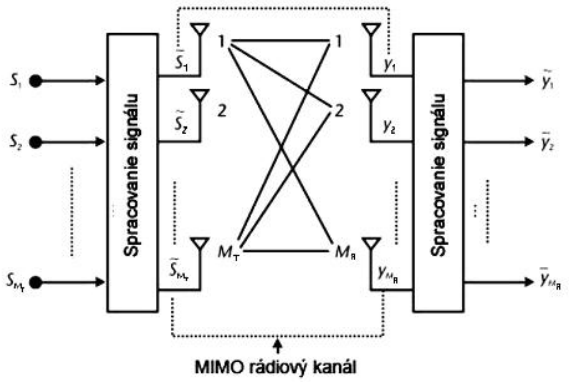
OFDM je širokopásmová modulácia využívajúca frekvenčné delenie kanála. Pracuje s tzv. rozprestretým spektrom, kedy je signál vysielaný na nezávislých frekvenciách. Následkom toho je zvýšenie odolnosti voči interferencií. Modulačná metóda OFDM vyžíva niekoľko stoviek až tisíc nosných frekvencií. Nosné sú ďalej modulované podľa potreby rôznymi moduláciami QPSK, 16-QAM alebo 64-QAM. Jednotlivé nosné frekvencie sú vzájomné ortogonálne, takže maximum každej nosnej by sa malo prekrývať s minimom ostatných. Vďaka tomu spektrum každej nosnej má nulovú hodnotu v strede pásma ostatných nosných frekvencií. Dátový tok celého kanála sa tak delí na stovky menších dátových tokov jednotlivých nosných frekvencií. Ak sú vo výsledku malé toky jednotlivých nosných, potom je možné vložiť ochranný interval GI (Guard-Interval) – je to čas kedy sa nevysiela žiadna nová informácia. Na prijímacej strane je tak možné bezproblémové prijatie (práve) vysielaného symbolu, aj keď prichádza k prijímaču viacerými cestami s rôznym oneskorením. Medzi typické MIMO techniky patrí priestorové multiplexovanie a priestorovo-časové kódovanie, ktoré sa môžu ľahko aplikovať na fyzickej vrstve OFDM komunikačných štandardov.

IV. MODEL MIMO SYSTÉMU

Predpokladáme bezdrôtový komunikačný MIMO systém s M_T vysielacími anténami a M_R prijímacími anténami (Obr. 2). V každom časovom okamihu sú signály S_i vysielané z antén súčasne. Každý vstup do prijímacích antén je lineárnou kombináciou prenášaných signálov, ktoré sú rušené šumom (jedna vysielacia a jedna prijímacia anténa predstavuje jednu cestu signálu). Na základe tohto modelu pre ľubovoľný signál z prijímacích antény potom platí [3]

$$y_{t,i} = \sum_{j=1}^{M_R} h_{i,j}.s_{t,i} + n_{t,j} \tag{1}$$

kde $n_{t,j}$ šumová vzorka prijímacej antény M_i v čase t. Aj keď sa rôzne verzie signálov s únikmi sčítajú na vstupe každej prijímacej antény, prítomnosť M_R kópii vysielaných signálov vytvára možnosť diverzitného zisku.



Obr. 2 Model MIMO kanála [19]

Stĺpcový vektor **S** tvorí vysielaciu maticu M_T x1, kde prvok S_i je i-tý prvok vyslaný i-tou anténou. Signály vysielané každou anténou majú rovnakú energiu rovnú E_S/ M_T, ak vysielač nepozná vlastnosti prenosového kanála. Anténny systém je realizovaný vysielacími a prijímacími anténami, ktoré definujú kanálovú maticu. Kanálová matica popisuje prenos signálu po rôznych cestách v rovnakom časovo frekvenčnom priestore a to bez zvýšenia spotreby výkonu. Za predpokladu kvázistacinárneho kanála pri prenosu jedného bloku dát, keď sa útlm nemení, bude mať matica kanála nasledovný tvar

$$\mathbf{H} = \begin{matrix} \\ M_R \\ \\ \end{matrix} \overbrace{\begin{bmatrix} h_{11} & h_{21} & \cdots & h_{M1} \\ h_{12} & h_{22} & \cdots & h_{M2} \\ \cdot & \cdot & \cdots & \cdot \\ h_{1M} & h_{2M} & \cdots & h_{MM} \end{bmatrix}}^{M_T}, \tag{2}$$

kde koeficienty h_{ij} predstavujú zisky ciest medzi vysielacou anténou M_i a prijímacou anténou M_j. Dá odvodiť aj maticová forma zápisu pre prijímané signály:

$$\boldsymbol{y} = \boldsymbol{H}\,\boldsymbol{s} + \boldsymbol{n}\,, \tag{3}$$

kde $\boldsymbol{s}$ je vektor vysielaných signálov, $\boldsymbol{y}$ vektor prijímaných signálov a $\boldsymbol{n}$ je vektor šumov. Ak nie sú antény od seba dostatočne vzdialené, je možné, že existuje určitá priestorová korelácia medzi ziskami ciest. Ak je vzdialenosť medzi dvoma anténami väčšia ako vlnová dĺžka, potom sú tieto zisky nekorelované [7], [9].

Všetky návrhy systémov sa snažia dosiahnuť určitú kapacitu priepustnosti. Pre SISO kanáli (Single-in-Single-out) sa kapacita počíta pomocou známeho Shannonovho vzťahu. Shannon definuje kapacitu ergodického kanála danú množstvom informácie, ktorá môže byť prenášaná s asymptoticky malou pravdepodobnosťou chyby. *Kapacita* takého kanála C je daná

v jednotkách bit/s alebo normalizovaním so šírkou pásma pomocou bit/s/Hz. Táto formulácia sa používa častejšie, nakoľko umožňuje jednoduchšie porovnávanie. Je zároveň nezávislá na šírke pásma.

$$C = log_2 (1+SNR) \quad b/s/Hz \tag{4}$$

Pri vysokých SNR (pomer signál/šum) sa ignoruje pričítanie 1 k SNR. Kapacita je potom priama funkcia SNR, teda $C = log_2 SNR$. Táto kapacita je založená na konštantnej prenosovej rýchlosti a nie je funkciou toho, či je informácia o stave kanála dostupná prijímaču alebo vysielaču. Tento výsledok je aplikovateľný iba na ergodické kanály, teda tie, u ktorých je ustálená prenosová rýchlosť a stabilné SNR. *Kapacita MIMO kanála* sa zvyšuje lineárne s počtom antén, pričom SISO/SIMO/MISO systémy zvyšujú kapacitu kanála iba logaritmicky (ide o obdivuhodné zvýšenie kapacity MIMO bez akéhokoľvek zvyšovania vysielacieho výkonu, iba pomocou zvýšenia počtu vysielačov). Štatistické vlastnosti kanála a korelácia medzi signálmi jednotlivých antén ovplyvňujú kapacitu MIMO kanála. Na zmenu kapacity hlavne vplýva korelácia. Predpokladáme, že iba prijímač nepozná prenosové vlastnosti kanála, kapacita pre N vysielacích a M prijímacích antén je vyjadrená vzťahom [4]

$$C_{EP} = log_2\left[det\left(I + \frac{\gamma}{N} H * H \right)\right] \quad b/s/Hz, \tag{5}$$

kde (*) vyjadruje transponovanú konjungciu, γ označuje SNR na prijímacej anténe a index EP definuje vysielače, ktoré vysielajú rovnaký výkon. Štúdie ukázali, že kapacita sa zvyšuje lineárne s min {M,N}. Vzťah (5) sa používa hlavne pre lineárny nárast kapacity Rayleigho kanála a nemusí platiť pre iné podmienky.

V. Záver

V článku je uvedený popis a analýza MIMO systému, ktorý sa začína využívať v bezdrôtovej komunikácií v automobilovej doprave. Aplikovaním systému dochádza k zvýšeniu dátovej priepustnosti, výkonu, dosahu vysielania, zároveň zvýšeniu počtu užívateľov bez rozšírenia šírky pásma na základe väčšej spektrálnej účinnosti. Veľmi dôležitým faktorom je typ prostredia, v ktorom sa prenášajú dáta. Na základe toho vieme určiť štandard, šírku pásma, rozsah nosných frekvencií. Tieto faktory ovplyvňujú úspešnosť použitia MIMO systému.

Ako prax ukazuje, testovanie MIMO systémov na vozovke je ekonomicky a technicky náročné, preto vhodnejším riešením je simulácia v simulačnom prostredí. Medzi najznámejšie spoločnosti, ktoré takéto riešenie podporujú patrí Altera, Mathworks, Texas Instruments a Xilinx. Napriek tomu je potrebné pri návrhu systémov, ako aj anténnych konfigurácií, prihliadať aj na testovanie v reálnych podmienkach z dôvodu vplyvu prostredia na prenášaný signál. Medzi technológie, ktoré využívajú MIMO systém pri prenose signálu v automobilovej doprave patrí aj Wi-Fi aj WiMAX.

Článok sumarizuje použitie MIMO systému v najbežnejších technológiách, ktoré majú veľký potenciál pre ďalší rozvoj. Svetové spoločnosti si tiež uvedomujú potrebu bezpečnejšej premávky a preto začínajú aplikovať jednotlivé systémy pre komunikáciu v automobilovej doprave, ale zatiaľ problémom je to, že systémy často nie sú kompatibilné.

Poďakovanie

Podporujeme výskumné aktivity na Slovensku / Projekt je spolufinancovaný zo zdrojov EÚ. Tento článok bol vypracovaný v rámci projektu "Centrum excelentnosti integrovaného výskumu a využitia progresívnych materiálov a technológií v oblasti automobilovej elektroniky", ITMS 26220120055.

LITERATÚRA

[1] L. Reichardt – T. Schipper – T. Zwick, "*Virtual Drive" Physical Layer Simulations for Vehicle-to-Vehicle Communication"*. URSI International Symposium on Electromagnetic Theory, 2010, pp.883 - 886.
[2] O. Renaudin et al, "*Wideband MIMO Car-to-Car Radio Channel Measurements at 5.3 GHz*". IEEE 68thVehicular Technology Conference VTC 2008-Fall, 2008, pp.1 -5.
[3] T. Kaieser et al, "*Prototyping for MIMO systems: An Overview*". In Proc. 12th European Signal Processing Conference (EUSIPCO '04), Vienna, Austria, Sept. 2004, pp.681- 688.
[4] P. Drotár, "*MIMO systémy pre mobilné komunikačné systémy s mnohonásobnými nosnými frekvenciami"*. Písomná práca k dizertačnej skúške. TU Košice, 2008.
[5] J. Maurer, T. Fügen, T. Schäfer, and W. Wiesbeck, "*A New Inter-Vehicle Communications (IVC) Channel Model*". In Proc. of the 60th IEEE Vehicular Technology Conference VTC2004-Fall, Sept. 2004, pp. 9–13.
[6] M. Matthaiou, et al, "*Capacity Study of Vehicle-to-Roadside MIMO Channels with a Line-of-Sight Component*". Wireless Communications and Networking Conference, 2008, pp.775-779.
[7] D. Kornek et al, "*Effects of Antenna Characteristics and Placements on a Vehicle-to-Vehicle Channel Scenario"*. IEEE Int. Conf. on Communications Workshops (ICC), 2010, pp.1-5.
[8] L. Reichardt et al, "*Capacity Maximizing MIMO Antenna Design for Car-to-Car Communication*". International Workshop on Antenna Technology (iWAT), 2011, pp.243-246.
[9] A. Paier et al, "*Overview of Vehicle-to-Vehicle Radio Channel Measurements for Collision Avoidance Applications*". IEEE 71st Vehicular Technology Conference (VTC 2010-Spring), 2010, pp. 1-5.
[10] J. MIČEK, "*Komunikačný systém pre zvýšenie bezpečnosti a efektívnosti*". Automatizace, roč. 51, č.4, 2008.

Modified Simmons' Authentication Scheme Using Elliptic Curve Arithmetic and Combinatorial Design

Patrik GALLO, Dušan LEVICKÝ

Department of Electronics and Multimedia Communications, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Slovak Republic

patrik.gallo@tuke.sk, dusan.levicky@tuke.sk

***Abstract* — In this paper we present an effective mutual authentication based on authentication scheme proposed by Gustavus Simmons in conjunction with elliptic curve arithmetic used for stronger security requirements and effective model for identifying the source of the message. Alternatively, we can see this model as authentication using same level of security with fewer bits required for keys in public key operations, thus faster applications can be implement in whole process of communication. Our eyes are set on objectives that makes authentication more effective – ease of use and strong authentication model.**

***Keywords* — Elliptic curve arithmetic, authentication, Simmons's model**

I. INTRODUCTION

An authentication scheme is a way of obtaining authentication protection (identifying either a source of the message or message itself), while a secrecy scheme (or security model) is a way of obtaining secrecy protection – obtaining confidentiality. It is clear that authentication and confidentiality are two different issues of information security. Authentication model may be with or without secrecy. In many situations we need to accomplish both of them, secure communication between banks and the customer, confident communication in between corporations, internal corporate communication and so on.

The rapid progress of developing mathematical principles and models that are used in security applications increased some questions about usage of these models in practical world. A lot of them provide either high level of security with lower efficiency in some applications with regard of bit-rate, or good efficiency but rely on small key-sizes that have to be changed in short terms. Our proposal combines advantages of elliptic curve arithmetic and effective authentication method that carries security out from combinatorial design.

The Elliptic Curve Cryptosystem is an emerging alternative for traditional Public-Key Cryptosystem like RSA, DSA and DH. It provides the highest strength-per-bit security of any cryptosystem known today with smaller key sizes resulting in faster computations, lower power consumption and memory storage. It also provides a methodology for obtaining high-speed, efficient and scalable implementation of protocols for authentication.

II. AUTHENTICATION SCHEME

In 1949 Shannon [1] showed how to construct authentication and secrecy system offering unconditional authentication and secrecy, that is, theoretically perfect-secrecy system. By the term unconditional, Shannon considered system that its security is independent of the computational power that the enemy (intruder) may use. Following this work, Simmons [1] and others have considered systems which offer unconditional authentication. Both unconditional secrecy and unconditional authentication are achieved at the expense of requiring a very large number of keys. To reduce the number of required keys and hence to reduce the key sizes in these systems as much as possible is an essential issue to be considered.

We usually consider an authentication scheme that involves three participants: a transmitter, a receiver and an intruder (opponent). Communication conditions are very simple: the transmitter wants to send a message to the receiver through insecure channel as the intruder

wants to impersonate the communication. According to literature, the model with three participants is usually called an A-code (Authentication code) [2]. We will also stick to this notation as it is commonly used in many other publications, considering authentication. The very simple A-code and its description is presented in [2] with all of its characteristics and mathematical background. Another authentication code is so called A²-code which introduces an arbiter in the whole communication process. The reason why the arbiter comes to the process of authentication is that the two parties (the transmitter and the receiver) are assumed trusted, they do not cheat each other. However, this is not always the case: the transmitter can easily deny sending any message, and receiver can attribute a fraudulent message to the transmitter. Thus, arbiter comes in between communicating parties to decide and agree on encoding rule used in particular session.

III. Elliptic Curve Menezes-Qu-Vanstone (ECMQV) Key Agreement

Elliptic curve cryptography (ECC) is a public key encryption technique based on elliptic curve theory that can be used to create faster, smaller, and more efficient cryptographic keys. ECC generates keys through the properties of the elliptic curve equation instead of the traditional method of generation as the product of very large prime numbers [3]. ECC is based on properties of a particular type of equation created from the mathematical group (a set of values for which operations can be performed on any two members of the group to produce a third member) derived from points where the line intersects the axes. Equations based on elliptic curves have a characteristic that is very valuable for cryptography purposes: they are relatively easy to perform, and extremely difficult to reverse. There are a lot of algorithms that uses elliptic curve arithmetic defined above and several protocols that are standardized in several organizations like NIST, IEEE, etc. For our purpose we chose ECMQV (Elliptic Curve Menezes-Qu-Vanstone) key agreement algorithm for defining the plane of the authentication grid, which will be explained in the next section.

The transmitter generates a key pair (T,t) by generating randomly t and calculating $T=tP$ where P is a point on elliptic curve. Receiver generates a key pair (R, r) in the same way as the transmitter. Now, the transmitter calculates:

$$S_a = t + \overline{T}a$$

and sends T to the receiver. Receiver calculates:

$$S_b = r + \overline{R}b$$

and sends R to the transmitter. Transmitter and receiver calculate:

$$K = h.S_a(R + \overline{R}B)$$
$$K = h.S_b(T + \overline{T}A)$$

where (A,a) and (B,b) are the key pairs of transmitter and receiver; A,B – public keys; a,b – private keys. $\overline{T}$ has the following meaning: let $T=(x,y)$be a point on an elliptic curve. Then

$$\overline{T} = (x \bmod 2^{\lceil f/2 \rceil}) + 2^{\lceil f/2 \rceil}$$

and

$$f = \lfloor \log_2 n \rfloor + 1$$

f represents the bit length of n.

This algorithm can be viewed as an extension of the ordinary Diffie-Hellman key agreement protocol. The used public keys are either static or temporary. Temporary keys are not necessarily authenticated, so if authentication is needed, it has to be obtained by other means. Static public keys provide neither forward secrecy nor key-compromise impersonation resilience, among other advanced security properties [4].

IV. Combinatorial Design and Secrecy of the Proposed Model

A transmitter communicates a sequence of distinct source states (messages) from a set S to a receiver by encoding them using one encoding rule from a set of E *(Key1, Key2,…)*. By encoding rules we mean using one of the rows that represents *Keys*, to map messages from the

matrices to the codes *(Code, Code2, …)*. Key in this scheme is represented by the decomposition of the source states in the matrices. However, each key (that represents the line) in the encoding matrices is derived from the operation of key agreement based on ECMQV explained in previous section. For obtaining exactly *n* encoded messages and encoding rules we simply introduce the modular operation to the process. We can entitle *n* as a cardinality of the matrices. So every derived key from ECMQV process before it can be used is divided by *n* and the remainder of that operation is set as encoded rule.

Each encoding rule is an injective mapping from *S* to the set of codes *M*:

$$S \to M : \text{Code}x = e(m_y)$$

where

$$\text{Key}z \in E,\ \text{Code}x \in M, m_y \in S$$

and

$$x, y, z \in N.$$

Mapping function *e* defines an encoding rule. Encoding rule uses established key from ECMQV to encrypt messages in the matrices. Encryption may be used either symmetric (faster) or asymmetric using elliptic curve integrated encryption scheme *S, M, E* is the set of source states, encoded messages and the set of encoding rules, respectively. Simple example of the encoding matrices is showed in Fig.1.

	Code 1	Code 2	Code 3	. . .	Code n
Key 1	m_1	—	m_2		—
Key 2	—	m_1	m_3		m_i
Key 3	—	m_3	—		m_{i-x}
⋮					
Key n	m_3	m_2	—		m_{i-2x}

Fig. 1 Encoding matrices

When the transmitter wants to send a message to the receiver, he simply encode it (find adequate encoded message “Code” in the matrices) by replacing it with “Code”. Assume that the probability of the given key is 1/*n*, because of the condition that we trying to propose an unconditional authentication scheme. There are exactly *i* possible encoded messages for each encoding rule, so the all others are invalid. So whenever somebody picks an encoded message in some particular key row from the column where there is no valid message *m*, authentication is rejected. That’s a first strong security issue in this model. When the intruder chooses an impersonation attack, he picks up an encoded message, “Code” and sends it to the receiver. The probability that particular “Code” will be accepted by the receiver is 1/*i*, for the assigned “Key” used at the moment. When the intruder launches a substitution attack, he embeds the encoded message sent by the transmitter and replaces it by a fraudulent one. He observed “Code”, thus he knows that the particular “Key” was used as it is shown on the Fig.1. So, other possibilities that he can choose for substitution attack are those that match the “Key” line with the valid messages in the line for this key. Then he picks up one of them as his encoded message and sends it to the receiver. His successful probability is also 1/*i*. Hence, the scheme from Fig.1 has the authentication function. And as the messages in each row for some “Code” but different “Keys” are not same, it also has the desired secrecy, since each encoded message can be used to transmit each message from the matrices. Intruder can just guess which message was transmitted with probability 1/*i*. To reach unconditional property of the model in sense of strong authentication, another is required.

The probability for occurrence of messages from the set *S* is denoted as *p(s)*, similarly for *p(e), p(m)* where *e* belongs to *E* and *m* to *M*. We assume that probabilities are independent, hence:

$$p(s,e) = p(s)\,p(e)$$

For any message $m \in M$,

$$p(m) = \sum p(s)\,p(e)$$

where the sum goes over all pairs *(s, e)* such that *e(s)=m.* We also require for every encoding rule, source state and encoded message that

$$p(e) > 0,\ \ p(s) > 0,\ \ p(m) > 0$$

Another useful measurement of unconditionality of proposed authentication scheme is the *entropy H*. It is known that entropy is a measurement of indeterminacy of some random variable [1]. As we stated that two probabilities *p(s)* and *p(e)* are independent, the entropy of occurrence any element from *S* is:

$$\begin{aligned} H(S \mid E) &= -\sum_{s,e} p(s,e)\log p(s) \\ &= -\sum_{s} \log p(s) \sum_{e}\ p(s,e) \\ &= -\sum_{s} p(s)\log p(s) = H(S) \end{aligned}$$

The important fact of this model is that the bigger the source state is the bigger entropy we get. Since we assumed that probabilities are always independent to each other and probabilities of particular source messages, encoded messages and encoding rules are equal, we obtain maximal entropy.

V. ELLIPTIC CURVE BACKGROUND OF THE PROPOSED MODEL

Before the exact usage of the explained authentication model one needs to generate the elliptic curve which is used for key agreement and assigning the encoding rules in the matrices. The whole process of generating the elliptic curve regarding all security requirements is described in [6].

The desired elliptic curve over finite field F_p is of the form:

$$C: y^2 = x^3 + ax + b\ ,$$

where all coefficients $a, b \in F_p$.

There is very important parameter that determines the elliptic curve and its singularity – *SeedC.*

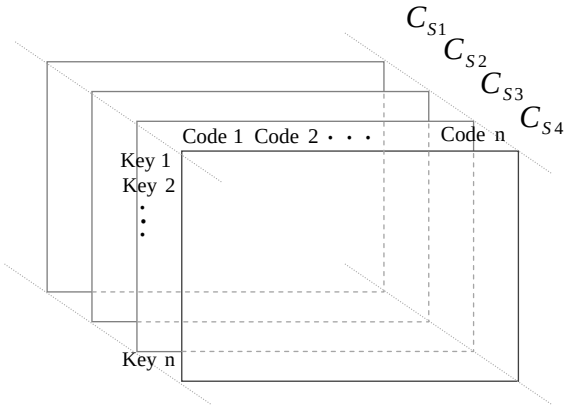


Fig. 2 Grid of matrices

Two kinds of curves can be given:

A. *Pseudo-random curves*

Pseudo-random curves are those whose coefficients (parameters needed to generate the curve) are generated from the output of a seeded cryptographic hash. If the seed value (*SeedC*) is given along with the coefficients, it can be verified easily that the coefficients were indeed generated by that method.

B. Special curves

Special curves whose coefficients and underlying field have been selected to optimize the efficiency of the elliptic curve operations. These special curves can be easily employed in the process of generating curves and consequently the point on the elliptic curve. If different seeds are specified, different elliptic curves are generated, thus different points on elliptic curves may be established.

This implies possibility of using several authentication matrices for different purposes regarding different security requirements. Figure 2. shows an example of a grid made of 4 authentication matrices (as mentioned above, possibly with unequal security property). These 4 matrices, represented by their elliptic curves C_{SX} generated separately by their own *Seeds* can be even used in authentication process by the same users for different purposes. Reason of this is that authentication process is very easy itself, only generation of elliptic curve and finding point on curve is little bit more complex. Although, using projective coordinates in process of computing points, makes this task easier and more effective [3].

VI. CONCLUSION

This paper studies authentication scheme with regards to combinatorial principles and elliptic curve cryptography. Proposed scheme uses Simmons' authentication scheme in conjunction of effective and secure ECMQV key agreement to deploy fast and strong authentication model. There are several advantages of this method. First is a simple structure of Simmons' method that provide fast encoding, as its simply uses matrices of size *n* to assigning (encrypt) source states into coded messages. The bigger *n* we chose, the bigger security (indeterminacy) we get, as its enlarge the sets *S*, *E* and *M*.

Second, matrices initializing by ECMQV provide strong and secure key deriving as it is based on solving elliptic curve discrete logarithm problem, which is still solvable just in subexponencial time. That shortly means, defining the frequently changing keys is impossible.
Third, with different seeds we can generate several elliptic curves that can be assigned to different processes or communication sessions, which differs in the level of security and time consumption.

Bigger sets on the other hand imply bigger complexity of the authentication scheme, which will result in longer processing time. This scheme is mentioned to use in fast communication applications with high level of authentication security combining above mathematical principles.

ACKNOWLEDGEMENT

This work is a result of the project implementation Development of the Center of Information and Communication Technologies for Knowledge Systems (project number: 26220120030) supported by the Research & Development Operational Program funded by the ERDF.

REFERENCES

[1] SIMMONS, G.L. Authentication theory/coding theory. Advances in Cryptology – Crypto'84, Lecture Notes in Computer Science 196, Springer- Verlag, Berlin, 1985
[2] PEI, D. Authentication codes and combinatorial design (Book style), Taylor & Francis Group, 2006
[3] MARTIN, L. Introduction to identity-based encryption (Book style), Norwood, MA: Artech house, 2008.
[4] HANKERSON, D, MENEZES, A, VANSTONE, S, (Hardcover): Guide to Elliptic Curve Cryptography, Springer Professional Computing, 2004
[5] STALLING, W. Cryptography and network security (Book style) Prentice Hall; 5 edition, January 2010.
[6] NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY, Recommended elliptic curves for federal Government use, July 1999

DCT vizuálneho objektu pomocou ortogonalizácie jej bázových funkcií v oblastiach hraničných blokov

Iveta GLADIŠOVÁ, Ján MIHALÍK

Katedra elektroniky a multimediálnych telekomunikácií, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

iveta.gladisova@tuke.sk, jan.mihalik@tuke.sk

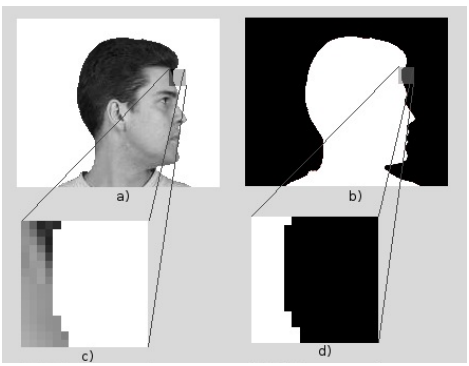
***Abstrakt* — V článku je popísaný princíp Gram-Schmidtov algoritmu ortogonalizácie bázových funkcií a jeho aplikácia pri diskrétnej kosínusovej transformácie vizuálneho objektu v oblastiach hraničných blokov. Získané výsledky touto metódou boli porovnávané s klasickou metódou diskrétnej kosínusovej transformácie vizuálneho objektu. V závere je vyhodnotenie výsledkov efektivity kódovania, pomerov signálu k šumu v hraničných blokoch obrazu vplyvom prahovej filtrácie transformačných koeficientov na kvalitu spätne transformovaných obrazov.**

***Kľúčové slová* — bázové funkcie, obrazový segment, Gram-Schmidtov algoritmus ortogonalizácie, dvojrozmerná transformácia obrazu**

I. ÚVOD

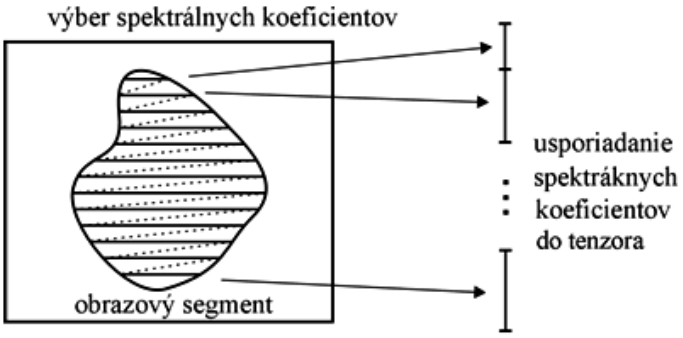
Nové algoritmy metódy kompresie obrazu pomocou výberu vizuálnych objektov z obrazovej scény vo všeobecnosti kódujú jednotlivé videoobjekty namiesto celých obrazových rámcov a podporujú tak rozvoj aplikácií s obsahovou interakciou. Z vizuálnej scény sa vytvárajú videoobjekty, ktoré sa môžu kódovať rôznymi technikami [1] - [3], [9]. Videoobjekt je zostavený z niekoľkých vrstiev za účelom jeho kódovania, ktorého syntax umožňuje rekonštrukciu videoobjektu zo základnej vrstvy a pridávanie ďalších doplnkových vrstiev za účelom zlepšenia jeho kvality.

V niektorých aplikáciách nie je potrebné rozlišovať videoobjekty kvôli zložitosti ich generovania a rozsiahlej prídavnej informácie. Vtedy sa kódujú pravouhlé snímky, ktoré predstavujú degenerované vyjadrenie tvaru videoobjektu [4], [5]. Pre účely transformácie a spracovania je obrazová rovina rozdelená do blokov obrazových prvkov (OP) s veľkosťou 4x4, 8x8 alebo 16x16. Aplikáciou binárneho tvaru na pôvodný obraz sa určuje, ktorý blok je vo vnútri, mimo, či na hranici videoobjektu. Pre každý z týchto prípadov je pripravené iné spracovanie bloku: a) bloky vo vnútri videoobjektu sa transformujú dvojrozmernou (2R) DCT, b) bloky mimo videoobjektu sa neprenášajú, ich ďalšie spracovanie je irelevantné, c) hraničné bloky (Obr. 1) sa transformujú segmentovo pomocou bázy 2R DCT ortogonalizovanej Gram-Schmidtovým algoritmom [6].



Obr.1 a) c) Výber hraničného bloku zo vstupného obrazu; b) d) výber hraničného bloku z binárneho tvaru obrazu.

Podľa toho, koľko je OP v danom segmente, sa budú vytvárať bázové funkcie. Ponorením segmentu ľubovoľného tvaru do pravouhlého okna (okná tvoria dvojrozmerné diskrétne kosínusové funkcie) sa vyberajú len tie spektrálne (transformačné) koeficienty, ktoré sú vo vnútri daného segmentu. Spektrálne koeficienty sa vyberajú po riadkoch od ľavého okraja segmentu po pravý okraj [3] a ukladajú sa do tenzora, ako je znázornené na obr. 2.



Obr.2 Výber spektrálnych koeficientov z obrazového segmentu a ich usporiadanie.

Takto získaný tenzor je vlastne bázová funkcia, pričom sa vytvorí len toľko bázových funkcií, koľko je OP v segmente. Tieto bázové funkcie nie sú ortogonálne, preto sa musia ortogonalizovať, a to napr. Gram-Schmidtovým algoritmom , ktorým sa budeme ďalej zaoberať.

II. 2. GRAM-SCHMIDTOV ALGORITMUS ORTOGONALIZÁCIE

Predtým, ako zistíme všeobecný Gram-Schmidtov algoritmus ortogonalizácie [7], budeme sa zaoberať problémom nájdenia vhodnej druhej bázovej funkcie (vektora), ktorá je ortogonálna k tej prvej. Začneme lineárne nezávislými vektormi $\bar{u}_0, \bar{u}_1, ..., \bar{u}_n$, kde n je počet OP v segmente. Vektor $\bar{q}_1$ sa určí podľa vzťahu

$$\bar{q}_1 = \bar{u}_1 - r_{01}\bar{n}_0 , \tag{1}$$

kde normovaný vektor

$$\bar{n}_0 = \frac{\bar{q}_0}{\left\|\bar{q}_0\right\|}, \qquad \text{a} \qquad \bar{q}_0 = \bar{u}_0 . \tag{2}$$

Parameter $r_{01} = \left\langle \bar{n}_0, \bar{u}_1 \right\rangle$ označuje skalárny súčin dvoch vektorov, potom po dosadení do rov. (1) dostaneme

$$\bar{q}_1 = \bar{u}_1 - \left\langle \bar{n}_0, \bar{u}_1 \right\rangle . \bar{n}_0 . \tag{3}$$

Platí, že $\bar{q}_1$ je ortogonálna k funkcii $\bar{q}_0$. Výraz $r_{01}.\bar{q}_0$ predstavuje projekciu vektora $\bar{u}_1$ na vektor $\bar{q}_0$. Test ortogonality:

$$\left\langle \bar{q}_1, \bar{n}_0 \right\rangle = \left\langle \left\langle \bar{u}_1 - r_{01}\,\bar{n}_0 \right\rangle, \bar{n}_0 \right\rangle = \left\langle \bar{u}_1, \bar{n}_0 \right\rangle - r_{01}\left\langle \bar{n}_0, \bar{n}_0 \right\rangle = 0 \tag{4}$$

Vo všeobecnosti môžeme písať:

$$\bar{q}_k = \bar{u}_k - r_{0k}\,\bar{n}_0 - r_{1k}\,\bar{n}_1 - - r_{2k}\,\bar{n}_2 - ... r_{k-1k}\,\bar{n}_{k-1} \tag{5}$$

pričom platí, že

$$r_{mk} = \left\langle \bar{n}_m, \bar{u}_k \right\rangle \qquad \text{a} \qquad \bar{n}_m = \frac{\bar{q}_m}{\left\|\bar{q}_m\right\|} \tag{6}$$

Použitím maticového zápisu dostaneme: matica daných bázových funkcií

$$\underline{U} = \left\{ \bar{u}_0, \bar{u}_1, ..., \bar{u}_n \right\}^T , \tag{7}$$

matica ortogonálnych bázových funkcií

$$\underline{Q} = \left\{ \bar{q}_0, \bar{q}_1, ..., \bar{q}_n \right\}^T . \tag{8}$$

Platí, že $\underline{U} = \underline{Q}.\underline{R}$, (9)

kde $\underline{R}$ je horná trojuholníková matica pozostávajúca z ortogonálnych koeficientov

$$\underline{R} = \begin{bmatrix} \frac{1}{\left\|\bar{q}_0\right\|} & r_{01} & r_{02} & ... & r_{0n} \\ 0 & \frac{1}{\left\|\bar{q}_1\right\|} & r_{12} & ... & r_{1n} \\ 0 & 0 & \frac{1}{\left\|\bar{q}_2\right\|} & ... & r_{2n} \\ . & . & . & ... & . \\ 0 & 0 & 0 & ... & \frac{1}{\left\|\bar{q}_n\right\|} \end{bmatrix}. \tag{10}$$

Hore uvedený algoritmus dáva ortogonálne bázové funkcie krok po kroku, čo znamená, že začíname s dvoma danými funkciami , z ktorých sa získajú dve ortogonálne funkcie. Postupne tretiu, štvrtú atď. Ortogonálna funkcia môže byť vypočítaná až pokiaľ $\underline{U}$ a $\underline{Q}$ nie sú štvorcové. To sa stane vtedy, ak máme toľko funkcií, koľko máme daných OP v segmente. Každá ďalšia funkcia musí byť lineárne závislá na ostatných [8], [10].

III. Experimentálne výsledky simulácie a ich zhodnotenie

Programové prostriedky na implementáciu Gram-Schmidtovho algoritmu ortogonalizácie bázových funkcií boli vytvorené v prostredí Matlab. Program je určený na výber ľubovoľného vstupného obrazu, z ktorého sa najprv vytvorí binárny tvar obrazu. Ten sa ďalej transformuje a to dvomi spôsobmi. *Hraničné bloky* sa transformujú segmentovo pomocou bázy 2R DCT ortogonalizovanej Gram-Schmidtovým algoritmom a *vnútorné bloky* sa transformujú klasickou 2R DCT. *Vonkajšie bloky* sa netransformujú.

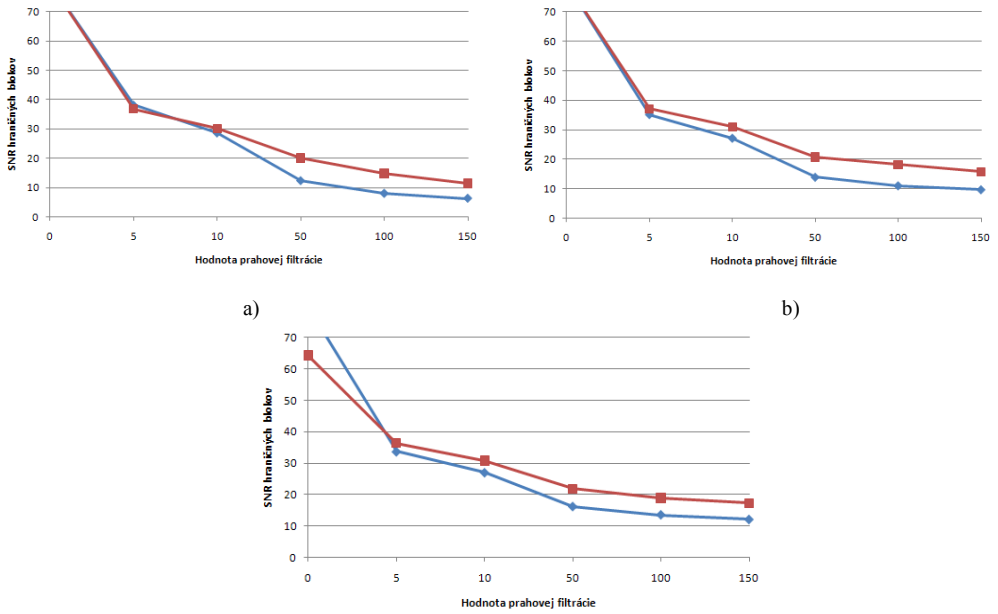
Vstupný obraz musí spĺňať tieto parametre: RGB formát obrazových dát, 8-bitové rozlíšenie jednej zložky obrazových dát RGB , počet riadkov a stĺpcov musí byť deliteľný 16, (ak nie je, program vypíše upozornenie). Vstupné obrazy musia byť uložené v jednom z týchto štandardných formátoch obrazu: jpg, bmp, png. Ak je načítaný farebný obraz, tak ho program prekonvertuje na obraz šedý. Obraz sa transformuje po blokoch štvorcových matíc, pričom je možnosť nastavenie voľby veľkosti týchto blokov na 4x4, 8x8, 16x16, tomu následne zodpovedá aj veľkosť jadra transformácie. V programe je možné vybrať farbu pozadia videoobjektu. Pri kompresii dát sa využíva skutočnosť, že DCT dekoreluje signál a koncentruje informačnú energiu do ľavej hornej časti matice transformačných koeficientov (koeficienty s najnižšími hodnotami energie je možné zanedbať, filtrovať, pričom nedochádza k veľkej degradácii obrazovej informácie). Program umožňuje *pri prahovej filtrácii* nastaviť prahovú hodnotu, ktorá vynuluje tie transformačné koeficienty, ktoré majú nižšiu hodnotu, ako je hodnota prahu. Po transformácii je možné navrhnutými programovými prostriedkami získať tiež informácie o počte hraničných blokov v celom obraze a o počte hraničných OP v týchto hraničných blokoch.

Zamerali sme sa na porovnanie účinnosti klasickej 2R DCT obrazu s modifikovanou metódou, ktorá transformuje hraničné bloky segmentovo pomocou 2R DCT ortogonalizovanej Gram-Schmidtovým algoritmom a vnútorné bloky transformuje klasickou 2R DCT. Vyhodnocovala sa miera kvality transformácie pomocou pomeru hodnoty signálu k šumu t.j. SNR (Signal to Noise Ratio). Sledovali sme tiež vplyv prahovej filtrácie na kvalitu spätne dekódovaného obrazu pre veľkosť blokov 4x4, 8x8 a 16x16 op (hodnota ponechaných

transformačných koeficientov bola väčšia, nanajvýš rovná prahovým hodnotám 0, 5, 10, 50, 100 a 150). Dosiahnuté výsledky pomeru SNR v dB v hraničných blokoch rôznych veľkostí sú uvedené v tab.1 a kvôli názornosti aj na obr.3.

TABUĽKA 1
POROVNANIE SNR HRANIČNÝCH BLOKOV PRE RÔZNE VEĽKOSTI BLOKOV

Prahová hodnota filtrácie	Blok 4x4		Blok 8x8		Blok 16x16	
	SNR [dB] klasická DCT	SNR [dB] Gram-Schmidt algorit. DCT	SNR [dB] klasická DCT	SNR [dB] Gram-Schmidt algorit. DCT	SNR [dB] klasická DCT	SNR [dB] Gram-Schmidt algorit. DCT
0	nekonečno	nekonečno	nekonečno	nekonečno	nekonečno	64,32
5	38,14	36,84	35,17	37,09	33,7	36,37
10	28,67	30,15	27,15	30,98	26,99	30,77
50	12,33	20,1	13,91	20,69	16,18	21,79
100	7,99	14,85	11,02	18,16	13,53	18,94
150	6,3	11,39	9,67	15,73	12,03	17,29

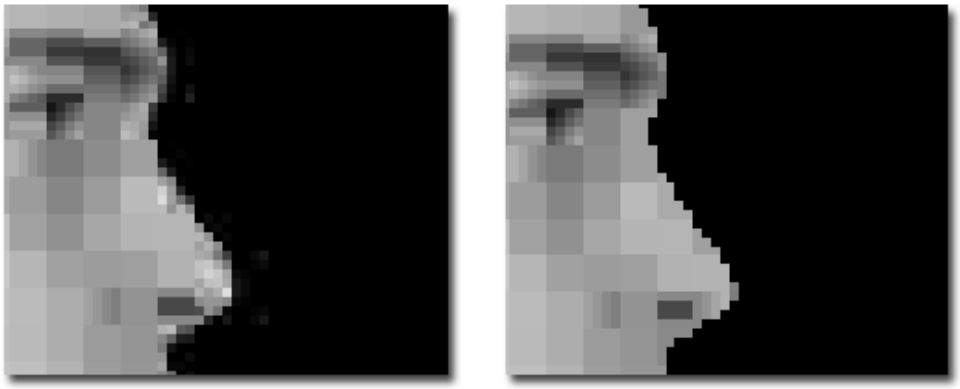


Obr. 3 Graf závislosti SNR hraničných blokov od hodnoty prahovej filtrácie *pri veľkosti bloku* a) 4x4 , b) 8x8, c) 16x16 (modrá – SNR v dB hraničných blokov pri klasickej DCT a červená (štvorce) – SNR v dB hraničných blokov pri DCT s ortogonalizáciou bázových funkcií)

Z dosiahnutých výsledkov vyplýva, že so vzrastajúcou hodnotou prahovej filtrácie klesá podľa očakávania pomer SNR hraničných blokov a zhoršuje sa kvalita obrazu. Pri nulovej hodnote prahovej filtrácie vychádzala hodnota SNR nekonečná (okrem bloku 16x16), čo znamená, že nedošlo k strate informácie pretože sa nefiltroval žiaden spektrálny koeficient. Pri bloku 16x16 hodnota SNR nebola nekonečná a to z dôvodu, že pri niektorých blokoch došlo k strate ortogonality a tým k malej chybe, aj keď bola nastavená nulová hodnote prahovej filtrácie. Pri blokoch 4x4 a 8x8 nedochádzalo k žiadnej chybe a táto metóda dosahovala najlepšie výsledky.

Výhodou modifikovanej metódy využívajúcej ortogonalizáciu bázových funkcií 2R DCT Gram-Schmidtovým algoritmom je, že vykazuje takmer vo všetkých prípadoch lepší pomer SNR ako klasická 2R DCT pri spracovaní hraničných blokoch (červená čiara na obr.3). Hodnota, o ktorú bol pomer SNR väčší sa pohybovala pri nízkych prahových hodnotách (5 resp. 10) od 1,48 dB po 3, 83 dB. Pri vyšších hodnotách prahu filtra sa zlepšenie SNR modifikovanej metódy pohybovalo od 5,09 dB po 7,77 dB, pričom najväčšie rozdiely boli pri veľkosti bloku 8x8. Treba si uvedomiť, že kvalita spätne transformovaného obrazu sa zhoršuje len v oblasti segmentu a nie celého bloku, čo má za následok zlepšenie kvality obrazu pri okrajoch objektu (obr.4). Vizuálny pocit pre pozorovateľa je preto oveľa lepší.

Nevýhodou takejto transformácie je v niektorých prípadoch nestabilita Gram-Schmidtového algoritmu, najmä pri veľkých veľkostiach blokov a tým vytváranie neortogonálnej bázy. Nevýhodou je aj časová náročnosť, ktorá so zväčšujúcim sa blokom rastie.



a) b)

Obr. 4Detail obrazu po spätnej transformácií pri hodnote prahovej filtrácie 50 a veľkosti bloku 4x4 a) klasická 2R DCT b) Gram-Schmidtovým algoritmom ortogonalizovaná 2R DCT

IV. ZÁVER

Článok bol zameraný na popis a simuláciu ortogonalizácie DCT bázových funkcii vizuálnych segmentov. Táto segmentová transformácia s ortogonalizovanou bázou DCT sa aplikovala na hraničné bloky vizuálneho objektu a v spojení s klasickou 2R DCT sa použila na transformovanie obrazu. Táto metóda sa porovnávala s klasickou blokovou 2R DCT obrazu pomocou dosiahnutej hodnoty pomeru signálu ku kvantizačnému šumu hraničných blokov pri rôznej veľkosti prahovej filtrácie a rôznych veľkostiach blokov. Zistilo sa, že táto metóda dosahovala oveľa lepší pomer signálu k šumu ako klasická 2R DCT. Z toho vyplýva, že táto metóda je vhodnejšia na spracovanie objektu a pri kompresií údajov vykazuje menšiu chybu pri spätnej transformácií. Nedochádza k takému veľkému skresleniu na hranici medzi objektom a pozadím ako pri klasickej 2R DCT.

Z výsledkov simulácií aj výpočtov vyplynulo, že Gram-Schmidtov algoritmus je niekedy nestabilný. Dochádza k strate ortogonality medzi vektormi alebo k vytvoreniu bázy vektorov, ktorá nie je ortogonálna od začiatku. Tým, že tento algoritmus používame na segmenty v obraze, ktoré sa nachádzajú v blokoch (čiže segment má menej op ako blok), vieme jeho nestabilitu ošetriť tak, že keď dôjde k strate ortogonality medzi bázovými funkciami máme postačujúce množstvo ďalších neortogonálnych bázových funkcií, ktorými vieme túto stratu napraviť a spustiť proces ich ortogonalizácie.

Segmentovo orientovaná transformácia DCT s ortogonalizáciou bázových funkcií má veľký potenciál a určite dosahujú sa ňou lepšie výsledky, ako pri blokovo orientovanej DCT, avšak za cenu zložitosti aj časovej náročnosti na výpočet.

LITERATÚRA

[1] Correia, P. - Pereira, F., *Video Analysis for Coding: Objectives, Features and Methods*, April 1997.
[2] Correia, P. - Pereira, F., *Coding Video Objects with the Emerging MPEG-4 Standard*. May 1997.
[3] Mihalík, J., Kódovanie *obrazu vo videokomunikáciách*. TU-FEI Košice, 2001.
[4] Mihalík,J., *Štandardný videokodek MPEG-4*. Slaboproudý obzor, roč. 60, č.2, 2003, s.7-11.
[5] Mihalík,J., *Štandardný videokodek H.264/AVC*, Slaboproudý obzor, roč. 67, č.4, 2011, str.4-10.
[6] Giraud, L. – Langou, J. – Rozlonik, M., *The Loss of Orthogonality in the Gram-Schmidt Orthogonalization Process*. Int. Journal of Computers and Mathematics with Applications, Vol.50, 2005, pp.1069-1075.
[7] Persson P., *Lecture 5: Gram-Schmidtov Orthogonalization*. Sept. 2007. http://persson.berkeley.edu/18.335/lec5handout6pp.pdf
[8] Ghodsi,S.R.- Mehr,B. - Taeibi-Rahn, M., *A Parallel implementation of Gram-Schmidt Algorithm for Dense Linear System of Equations*. International Journal of Computer Applications,Vol.5, No.7, August 2010, pp.16-20.
[9] R.C. Gonzalez, Digital image processing, 3rd ed., N.J., Prentice Hall, 2008.
[10] Ramesha K. - Raja, K.B.: Gram-Schmidt Orthogonalization Based Face Recognition using DWT. Int. Journal of Engineering Science and Technology (IJEST), Vol. 3, No. 1, 2011, pp.494-503.

Implementácia algoritmu mriežkového vektorového kvantovania

Iveta GLADIŠOVÁ

Katedra elektroniky a multimediálnych telekomunikácií, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

iveta.gladisova@tuke.sk

***Abstrakt* — Článok sa zaoberá problematikou mriežkového vektorového kvantovania a jeho rýchleho algoritmu pre mriežku bodov typu Dv. Následne poskytuje dosiahnuté výsledky simulácie navrhnutého algoritmu takéhoto mriežkového vektorového kvantizátora pre kvantovanie pseudonáhodných testovacích postupností vektorov so špecifikovanými štatistickými vlastnosťami a rozmeru 2 až 5. Experimentálne výsledky potvrdili, že zo vzrastajúcou sa prenosovou rýchlosťou údajov sa zlepšuje S/Š a tento pomer sa zlepšil aj pri zväčšovaní rozmeru. Pre každú prenosovú rýchlosť sa našli optimálne hodnoty ohraničenia množiny bodov mriežky, pre ktoré je S/Š mriežkového vektorového kvantizátora najlepší.**

***Kľúčové slová* — vektorové kvantovanie, mriežka bodov, rýchly algoritmus kvantovania**

I. ÚVOD

Iteratívny spôsob návrhu vektorového kvantizátora umožňuje návrh optimálneho vektorového kvantizátora. Táto metóda má však dva veľké nedostatky. Prvým nedostatok je, že vstupná cvičná postupnosť využíva iteračnú štatistickú metódu, ktorá vyžaduje dlhý čas výpočtu. Druhý nedostatok je významnejší. Predložený návrh nemá zovšeobecnenú štruktúru a implementácia návrhu optimálneho vektorového kvantizátora vyžaduje zložitý výpočet medzi vstupným a každým výstupným vektorom tak, aby mohol byť vybraný najlepšie vyhovujúci výstupný vektor. Všeobecne zložitosť návrhu tohto kvantizátora a zložitosť výberu kvantizačných vektorov rastie exponenciálne s prenosovou rýchlosťou údajov a s rozmerom vektora. Preto je samozrejmý záujem o vývoj takých spôsobov návrhu vektorových kvantizátorov, ktorých rast zložitosti výpočtu nemá exponenciálny charakter v závislosti od výslednej prenosovej rýchlosti údajov a rozmeru vektora.

Mriežkové vektorové kvantovanie (MVK) [1], [10], [11] je jednoduchým, ale silným nástrojom pre vektorové kvantovanie a môže byť vnímané ako vektorové zovšeobecnenie rovnomerného skalárneho kvantovania. Použitie mriežok pri vektorovom kvantovaní predurčuje symetrická geometrická štruktúra mriežky bodov (vektorov) spojená s jej pravidelnosťou a tiež rýchly algoritmus kvantovania, ktorý takáto štruktúra umožňuje [5]. V mriežkovom vektorovom kvantovaní je množina kvantizačných vektorov konštruovaná z konečnej podmnožiny mriežkových bodov. Symetria regulárnej mriežky robí kvantovanie a dekódovanie veľmi jednoduchým, ako výber a indexovanie najbližšieho bodu mriežky, t.j. kvantizačného vektora k danému vstupnému vektoru . Naviac, kvantizačné vektory nie je nutné uchovávať, môžu byť generované súborom jednoduchých algebrických pravidiel. Mriežkové vektorové kvantovanie ponúka veľkú redukciu zložitosti v procese vektorového kvantovania a eliminuje pamäťové požiadavky[6], [7], čo sú dva sprievodné problémy pri realizácii optimálnych vektorových kvantizátorov .

II. VEKTOROVÉ KVANTOVANIE NA BÁZE MRIEŽOK

Mriežkový vektorový kvantizátor je suboptimálny vektorový kvantizátor, ktorého kvantizačné vektory sú vektormi mriežky a je určený geometrickou štruktúrou mriežky. Vo všeobecnosti je v-rozmerná mriežka L_v reprezentovaná množinou bodov, resp. vektorov [2] - [4] ako

$$L_v = \left\{ \bar{b} \in R_k : \bar{b} = u_1 \bar{g}_1 + u_2 \bar{g}_2 + ... + u_v \bar{g}_v \right\} \quad (1)$$

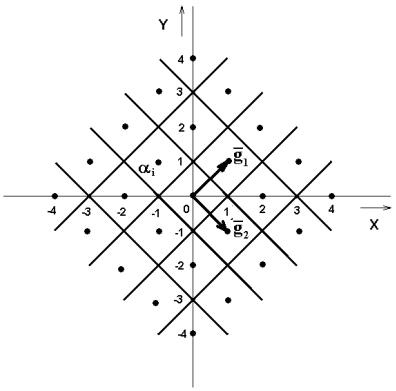
v k-rozmernom euklidovskom priestore R_k, kde $v \leq k$, ďalej u_1, u_2, ..., u_v sú celé čísla

a $\bar{g}_1, \bar{g}_2, ..., \bar{g}_v$ sú základné vektory (báza) mriežky.

Mriežka typu D_v pre rozmer $v \geq 2$, ktorá bola použitá pri simuláciách, je reprezentovaná množinou vektorov s celočíselnými zložkami, ktorých súčet je párne číslo, teda

$$D_v = \left\{ \bar{b} = (b_1, b_2, ..., b_v), \sum_{i=1}^{v} b_i = párne \quad číslo \right\}. \tag{2}$$

Pre rozmer v = 3 je D_3 plošne centrovaná kubická mriežka a najznámejšou duálnou mriežkou je D_3*, ktorá sa nazýva priestorovo-centrovaná kubická mriežka. Bázové vektory duálnej mriežky sú jednoducho reciproké bázové vektory regulárnej mriežky. Na Obr.1 je znázornená mriežka typu D_2 v dvojrozmernom priestore R_2 (v = 2) s vyznačenými bázovými vektormi $\bar{g}_1, \bar{g}_2$ a kvantizačnými (tzv. Voronoiho) oblasťami α_i [3].



Obr. 1 Mriežka typu D_2 v dvojrozmernom priestore R_2.

Pri návrhu MVK sú teda vektory mriežky kvantizačnými vektormi a kvantizačné oblasti sú vytvárané okolo týchto bodov mriežky, pričom kvantizačné vektory sú ťažiskami týchto oblastí (napr. pre celočíselnú mriežku Z_2 sú kvantizačné oblasti štvorce s kvantizačnými vektormi uprostred každého štvorca). Conway a Sloane [5] objavili rýchle a jednoduché algoritmy pre všetky známe mriežky do ôsmeho rozmeru. Ich algoritmy využívajú práve symetriu mriežok k nájdeniu najbližšieho mriežkového bodu pre daný vstupný vektor s minimálnym výpočtovým úsilím.

Pre zadanú prenosovú rýchlosť údajov je potrebné mriežky vhodne orezať (podľa počtu kvantizačných vektorov) a transformovať ich mierku t.j. škálovať. Nevyhnutné je poznať, koľko bodov mriežky leží vo vnútri orezanej plochy, t.j. nutné je vedieť tvar vymedzenej plochy. Platí, že pre zdroje s Laplaceovým rozdelením pravdepodobnosti sú vymedzenými plochami v-rozmerné pyramídy a pre zdroje s Gaussovým rozdelením pravdepodobnosti sú tieto plochy v-rozmerné gule. Polomer pyramídy alebo gule tak určuje tvar vymedzenej plochy mriežky s požadovaným počtom bodov, teda kvantizačných vektorov.

III. RÝCHLY ALGORITMUS KVANTOVANIA S MRIEŽKOU D_V

Pred popísaním rýchleho algoritmu mriežkového vektorového kvantovania pre mriežku D_v je potrebné zadefinovať niekoľko funkcií, ktoré zjednodušia jeho popis [5].

Nech pre reálne číslo *x* platí, že *f(x)* je najbližšie celé číslo ku *x*. Ak *x* je rovnako vzdialené od dvoch celých čísel, vezme sa celé číslo s menšou absolútnou hodnotou. Rozdiel medzi *x* a *f(x)* bude: $\delta(x) = x - f(x)$.

Funkcia *w(x)* bude predstavovať druhé najbližšie celé číslo ku *x*. V prípade rovnakej vzdialenosti od dvoch celých čísel, vezme sa číslo s väčšou absolútnou hodnotou. Pre ľubovoľný vektor $\bar{x} \in R_v$ sa budú vektorové verzie týchto funkcií nasledovné:

$$\bar{f}(x) = (f(x_1), f(x_2), ..., f(x_v)). \tag{3}$$

Pre dané $\bar{x}$, nech k je také celé číslo, že

$$|\delta(x_k)| \geq |\delta(x_i)|, \qquad 1 \leq i \leq v,$$

$$|\delta(x_k)| = |\delta(x_i)|, \qquad \text{ak } k = i \tag{4}$$

Funkcia $\bar{g}(x)$ bude mať tvar

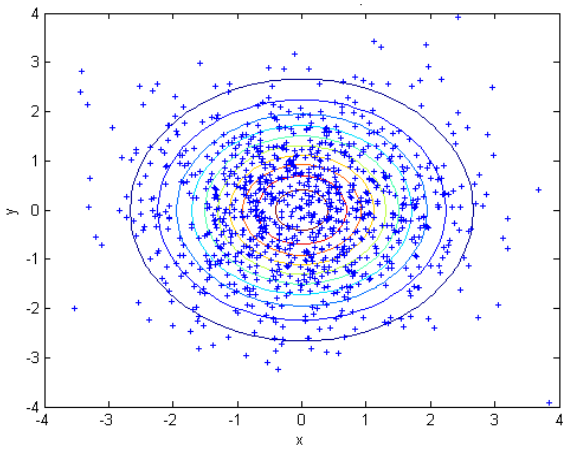
$$\bar{g}(x) = (f(x_1), f(x_2), ..., w(x_k), ..., f(x_v)). \tag{5}$$

Pri hľadaní najbližšieho vektora mriežky D_v k danému $\bar{x} \in R_v$ sa vychádza z toho, že najbližším vektorom mriežky D_v je ten z dvojice vektorov $\bar{f}(x)$ a $\bar{g}(x)$, ktorý má párny súčet zložiek svojho vektora. Ak súčet zložiek $\bar{f}(x)$ je nepárny, je garantované, že súčet zložiek $\bar{g}(x)$ je párny, a preto je to najbližší mriežkový bod, kvantizačný vektor. Vo všeobecnosti, ak vstupný vektor $\bar{x}$ je rovnako vzdialený od dvoch alebo viacerých vektorov mriežky, algoritmus vyberie vektor s najmenšou normou.

Algoritmus kvantovania pre mriežku Dv môže byť tiež základom pre algoritmy kvantovania iných optimálnych mriežok [8], [9], ako sú Gossetova E_8 a Barnesova-Wallova mriežka L_{16}, pretože tieto mriežky sú vytvorené zjednotením množín mriežok D_v.

IV. Dosiahnuté výsledky simulácie a ich zhodnotenie

Na základe rozpracovanej problematiky bol implementovaný algoritmus mriežkového vektorového kvantovania s mriežkou typu D_v pre rozmer v = 2 až 5 (od rozmeru 6 a viac sú iné typy optimálnych mriežok). Navrhnuté programové prostriedky boli testované pre cvičnú vstupnú postupnosť štatisticky nezávislých vstupných vzoriek s Gaussovým rozdelením pravdepodobnosti, s nulovou strednou hodnotou a jednotkovou disperziou. Z tejto postupnosti bolo možné vytvárať rovnako dlhé postupnosti vstupných vektorov pre požadované rozmery. Počet vstupných vektorov bolo možné nastaviť. Pred samotným kvantovaním bolo potrebné zadať požadovanú prenosovú rýchlosť údajov v bitoch na zložku, ktorá je závislá na počte kvantizačných vektorov MVK, ako aj na rozmere. Je potrebné si uvedomiť, že vo všeobecnosti sme pri počte kvantizačných vektorov limitovaní typom mriežky a jej ohraničením (tých bodov mriežky, ktoré sa stanú kvantizačnými vektormi). Preto sa pri simulácii MVK na základe požadovanej prenosovej rýchlosti urobila najprv ponuka možnosti jej výberu z troch najbližších možných hodnôt k požadovanej rýchlosti. Z tohto výberu potom už vyplynul definitívny počet kvantizačných vektorov MVK.



Obr. 2 Vektory z preťaženia v dvojrozmernom priestore R_2.

Vyhodnocoval sa dosiahnutý pomer signál ku šumu (S/Š) v decibeloch a tiež počet vektorov z preťaženia. Vstupné vektory, ktoré sa ocitli za hranicou ohraničenia množiny kvantizačných vektorov, sú považované za vektory z preťaženia a sú kvantované bodmi mriežky ležiacimi priamo na tejto hranici ohraničenia. So vzrastajúcou prenosovou rýchlosťou rastie aj počet kvantizačných vektorov a tým rastie aj polomer ohraničenia, z toho vyplýva pokles vektorov z preťaženia, čím sa zmenšuje kvantizačná chyba. Na Obr.2 je vidno rozmiestnenie vstupných dvojrozmerných vektorov s Gausovým rozdelením pravdepodobnosti a všetky tie, ktoré ležia za poslednou ohraničujúcou kružnicou, sú považované za vektory z preťaženia.

Rozhodujúcim parametrom bola prenosová rýchlosť, na základe ktorej sa porovnávali výsledné hodnoty. Prenosová rýchlosť sa zvyšovala po 0,5 bitov/zložku. Hrubým písmom sú v Tab. 1 až Tab. 4 vyznačené tie hodnoty celkového pomeru S/Š, pri ktorých je najmenší rozdiel medzi požadovanou a skutočnou (ponúkanou) hodnotou prenosovej rýchlosti údajov.

Z dosiahnutých výsledkov uvedených v týchto tabuľkách pre 10 000 vstupných testovacích vektorov rozmeru 2 až 5 vyplynulo, že:

- pri približne rovnakých prenosových rýchlostiach a rôznych rozmeroch nám so stúpajúcim rozmerom spracovania (rozmerom mriežky) rastie aj pomer S/Š;
- pri kvantovaní s menším počtom bitov/zložku (napr. 1,5) môžeme pozorovať vysokú mieru vektorov z preťaženia. Ak nechceme meniť prenosovú rýchlosť, tak počet vektorov z preťaženia je možné znížiť použitím metódy škálovania pre danú vstupnú postupnosť vektorov;
- ak si vyberieme jeden rozmer a sledujeme výsledky vo všetkých štyroch Tab.1 až Tab4. vidíme, že s rastúcou prenosovou rýchlosťou rastie aj pomer S/Š, pričom je možné pozorovať, že v určitom momente nastáva tzv. „stav nasýtenia“, kedy sa pomer S/Š už ďalej nezvyšuje a počet vektorov z preťaženia je nízky, resp. rovný nule. V takomto prípade je pre požiadavku zvýšenia pomeru S/Š výhodnejšie zväčšiť rozmer kvantizátora a teda aj rozmer vstupných vektorov (blokovo spracovať väčší počet vzoriek vstupného súboru), ako ďalej zvyšovať prenosovú rýchlosť.

Tabuľka 1
Porovnanie S/Š v [dB] MVK pri prenosovej rýchlosti 1,5 bit/zl

Požadovaná prenos. rýchlosť 1,5 [bit/zl]	MVK s D_2		MVK s D_3		MVK s D_4		MVK s D_5	
Ponúkaná prenos. rýchlosť [bit/zl]	1,16	**1,58**	**1,42**	1,81	**1,41**	1,79	**1,41**	1,71
Počet kvantizačných vektorov	5	9	19	43	49	145	131	371
Celkový pomer S/Š [dB]	7,39	**7,52**	**8,38**	9,12	**8,64**	9,42	**8,59**	9,47
Počet vektorov z preťaženia	3743	1339	2667	1133	4112	2117	5482	3038

Tabuľka 2
Porovnanie S/Š v [dB] MVK pri prenosovej rýchlosti 2 bit/zl

Požadovaná prenos. rýchlosť 2 [bit/zl]	MVK s D_2		MVK s D_3		MVK s D_4		MVK s D_5	
Ponúkaná prenos. rýchlosť [bit/zl]	**1,85**	2,2	**1,93**	2,10	1,85	**2,07**	1,83	**2,03**
Počet kvantizačných vektorov	13	21	55	79	169	313	571	1131
Celkový pomer S/Š [dB]	**8,23**	8,29	**9,26**	9,36	9,73	**9,89**	9,9	**10,08**
Počet vektorov z preťaženia	191	78	470	201	994	443	1538	765

Tabuľka 3
Porovnanie S/Š v [dB] MVK pri prenosovej rýchlosti 2,5 bit/zl

Požadovaná prenos. rýchlosť 2,5 [bit/zl]	MVK s D_2		MVK s D_3		MVK s D_4		MVK s D_5	
Ponúkaná prenos. rýchlosť [bit/zl]	**2,43**	2,60	**2,49**	2,55	2,47	**2,52**	**2,46**	2,55
Počet kvantizačných vektorov	29	37	177	201	937	1081	5033	6893
Celkový pomer S/Š [dB]	**8,3**	8,3	**9,41**	9,41	9,99	**9,99**	**10,25**	10,25
Počet vektorov z preťaženia	2	0	2	1	17	7	20	9

Tabuľka 4
Porovnanie S/Š v [dB] MVK pri prenosovej rýchlosti 3 bit/zl

Požadovaná prenos. rýchlosť 3 [bit/zl]	MVK s D_2		MVK s D_3		MVK s D_4		MVK s D_5	
Ponúkaná prenos. rýchlosť [bit/zl]	**2,96**	3,05	2,95	**3,02**	**2,99**	3,05	2,96	**3,00**
Počet kvantizačných vektorov	61	69	459	531	3961	4729	28863	33183
Celkový pomer S/Š [dB]	**8,3**	8,30	9,41	**9,41**	**10,00**	10	10,25	**10,25**
Počet vektorov z preťaženia	0	0	0	0	0	0	0	0

Nakoniec sa porovnával S/Š v dB pri mriežkovom vektorovom kvantovaní testovacích postupností s odlišnými hustotami rozdelenia pravdepodobnosti, a to Gaussove v porovnaní s Laplaceovým rozdelením pravdepodobnosti. Obe rozdelenia mali nulovú strednú hodnotu a jednotkovú disperziu, rovnaký počet vstupných vektorov. Porovnanie je urobené v Tab.5 pre 4-rozmerný MVK s mriežkou D_4 pri bitovej rýchlosti 1,5 a 2bit/zložku a pre 10 000 vstupných vektorov, bez škálovania.

Je zrejmé, že hodnoty prenosovej rýchlosti a hodnoty počtu kvantizačných rozmerov sú rovnaké pri oboch druhoch rozdelenia pravdepodobnosti. Menil sa iba dosiahnutý pomer S/Š a počet vektorov z preťaženia.

Tabuľka 5
Porovnanie S/Š v [dB] MVK pre Gaussove a Laplaceove rozdelenie pravdepodobnosti

Požadovaná prenos. rýchlosť [bit/zl]	1,5		2		1,5		2	
Ponúkaná prenos. rýchlosť [bit/zl]	**1,40**	1,79	1,85	**2,07**	**1,40**	1,79	1,85	**2,07**
Počet kvantizačných vektorov	49	145	169	313	49	145	169	313
Celkový pomer S/Š [dB]	**8,64**	9,42	9,73	**9,89**	**8,18**	8,97	9,34	**9,6**
Rozdelenie	Gaussove				Laplaceove			

Z porovnania vyplýva, že hodnota pomeru S/Š pri rovnakej prenosovej rýchlosti je v prípade Laplaceového rozdelenia pri oboch vybraných prenosových rýchlostiach nižšia v porovnaní s Gaussovým rozdelením pravdepodobnosti. Je to spôsobené tým, že vektory mriežky D_v, teda kvantizačné vektory, majú okrem iného aj takú vlastnosť, že ležia na v-rozmerných kružniciach a ohraničujúce oblasti MVK sú tiež kružnice, čo lepšie vyhovuje práve vstupným vektorom s Gaussovým rozdelením pravdepodobnosti. Pri Laplaceovom rozdelení pravdepodobnosti je potrebné mať iné ohraničenie oblastí výskytu vektorov (v-romerné pyramídy).

V. ZÁVER

Článok predstavil spôsob mriežkového vektorového kvantovania s použitím mriežky typu D_v pre rozmery vektorového priestoru 2 až 5, pričom sa využil rýchly algoritmus kvantovania pre daný typ mriežky. Uskutočnené boli simulácie tohto mnohorozmerného kvantovania pre testovaciu vstupnú postupnosť štatisticky nezávislých vektorov s konkrétnymi vlastnosťami. Získané výsledky simulácie poukazujú na to, že účinnosť MVK rastie s rozmerom pri tej istej prenosovej rýchlosti, ďalej S/Š rastie tiež so zvyšujúcou sa prenosovou rýchlosťou pre konkrétny rozmer. Výhoda MVK oproti optimálnemu vektorovému kvantovaniu je predovšetkým v jednoduchosti spracovania a v menšom počte operácii pri kvantovaní.

Mriežkový vektorový kvantizátor s mriežkou typu D_v môže byť použitý ako vhodná metóda kvantovania za účelom zvýšenia jeho výkonu. Tým je možné získať efektívny systém kódovania, ktorý môže byť dôležitou súčasťou digitálneho spracovania signálov z inteligentných senzorov, napríklad radarov alebo videosenzorov v oblasti automobilovej elektroniky a komunikačných systémov.

POĎAKOVANIE

Podporujeme výskumné aktivity na Slovensku / Projekt je spolufinancovaný zo zdrojov EÚ. Tento článok bol vypracovaný v rámci projektu "Centrum excelentnosti integrovaného výskumu a využitia progresívnych materiálov a technológií v oblasti automobilovej elektroniky", ITMS 26220120055.

LITERATÚRA

[1] F. Ch. Zheng Gao - J. Villasenor, "Lattice Vector Quantizer of Generalized Gaussian Sources". IEEE Trans. on Inform. Theory, vol.43, 1980, no.1, 1997, pp. 92-103.
[2] I. Gladišová – J. Mihalík, Mriežkový vektorový kvantizátor so sústrednými pyramídami a entropickým kódovaním. In: Electrical Engineering and Informatics: Proc. of the Faculty of Electrical Engineering and Informatics of the Technical University of Košice, September, 2010, pp.770-774..
[3] I. Gladišová – J. Mihalík, "Mriežkové vektorové kvantovanie s entropickým kódovaním. Zborník vedeckej konferencie s medzinárodnou účasťou " Nové smery v spracovaní signálov V" , L.Mikuláš, 2000, s.355-360.
[4] L.H. Fonteles – U. Antonini, "High Dimension Lattice Vector Quantizer Design for Generalized Gaussian Distributions". IEEE International Conference on Image Processing, Sept.-Oct.2007, vol.4, pp. IV-185 - IV-188.
[5] J. H. Conway and N. J. A. Sloane, "Fast quantizing and decoding algorithms for lattice quantizers and codes," IEEE Trans. Inform. Theory, vol. IT-28, Mar. 1982, pp. 227-232.
[6] M.F.M. Salleh – J. Soraghan, "A New Multistage Lattice Vector Quantization with Adaptive Subband Thresholding for Image Compression". EURASIP Journal on Advances in Signal Processing, Volume 2007, http://downloads.hindawi.com/journals/asp/2007/092928.pdf
[7] J. Ostergaard, "Multiple-Description Lattice Vector Quantization." PhD Thesis, 2007. http://homepage.tudelft.nl/c7c8y/Theses/PhDThesisOstergaard.pdf
[8] I. Syafalni – M.F.M. Salleh, "Multistage Lattice Vector Quantization with Adaptive Normalization for Wavelet-Based Image Coding", European Journal of Scientific Research, vol.41, no.1, 2010, pp. 48.-56.
[9] M.F.M., Salleh – J.Soraghan "A New Adaptive Subband Thresholding Algorithm for Multistage Lattice VQ Image Coding". ICASSP 2006 Proceedings, 2006, pp.457 -460.
[10] B. Hang - X. Wang - R. Hu - Y. Ma - Z. Xie, "An Even Grid Based Lattice Vector Quantization Algorithm for Mobile Audio Coding", Journal of Multimedia, vol. 6, no. 3, 2011, pp. 261
[11] Patchoo, Wisarn, "Analysis and application of lattice vector quantization using mixture models and bit-plane coding", PhD Thesis, 2011, http://gradworks.umi.com/34/79/3479195.html

Effective Measurement Stand for Teaching Electronic Circuits

[1]*Marek GODLA,* [1]*Linus MICHAELI,* [2]*Ján ŠALIGA,* [3]*Martin SEKERÁK,* [5]*Daniel FÁBRY*

[1] Department of Electronics and Multimedia Communications, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Slovak Republic
[2] Digi Slovakia, s.r.o., Einsteinova 21/3692 851 01 Bratislava, Slovak Republic

[1]marek.godla@tuke.sk, [1]linus.michaeli@tuke.sk,
[1]jan.saliga@tuke.sk,[1]martin.sekerak@tuke.sk, [2]dfabri@azet.sk

***Abstract*— the paper presents project and outcomes from remote access laboratory for experimental exercises. This project has been implemented in the course „Basics of electronics". The experimental laboratory stands are controlled remotely from client´s computers across the Internet. User can change configuration of circuit components, component values and also select measuring nodes. The control and data presentation is realized using virtual instruments implemented in dynamic web pages broadcasted by laboratory web server with installed plug-in "Run Time Engine" by National Instruments. This project has been created in programming environment LabVIEW. We have used for minimize realization cost DAQ cards instead of traditional stand-alone instrumentations.**

***Keywords* — remote access laboratory, basic of electronics, improve of teaching**

I. INTRODUCTION

The main task of experimental laboratory stands in the courses of electronic is obtaining practical and theoretical knowledge. Practical experiments allow students familiarizing with measuring instruments, assessment of accuracy of simulation results in comparison with measured ones and enhance the deepness and quality of achieved knowledge in the subject. For this solution it's possible to realize real measurement and verification of the acquired theoretical knowledge in laboratory. This approach requires the personal presence of a supervisor in the laboratory especially for the students starting the bachelor level of education and also restricts the laboratory access time according to university internal security and other conditions. The basis way when it's possible to get are two solution of this problem. One is virtual laboratory that is based on simulation. The main disadvantage of this approach is inaccuracy of the behaviour model of real measurement. The other way is used solution from remote laboratory based on real hardware. Design of remote laboratories for various application fields was studied by many authors, e.g., [1] – [3].

The aim of this paper is to describe hardware and software solution for teaching subject "Basic of Electronics" using real measuring workplace accessible across the Internet developed at the Department of Electronics and Multimedia Communications, Technical University of Kosice.

II. IMPLEMENTATION OF EXPERIMENTAL WORKS IN CURRICULA

The simple way in effective teaching is focused on electronic circuits at bachelor and master level of university education is not possible without simulation and real laboratory experiment.

It is very different between simulation experiment and real experiment in real laboratory measurement. Electronic circuit simulation is very simple way how it's possible to verify theoretic knowledge without risk to damage electronic components and event measuring instrumentation. Main knowledge is use simulation in big groups of students. It enables better understanding of relations among electronic components, their values, connections and the final properties of resulting electronic circuits.

On the other hand, the simulations sometimes leads students to the expectation that behaviour of real circuit is always in perfect equality to the results of simple simulation and that the

realized circuits will work exactly as it worked in simulation environment. This fact is the main reason why a real experimental measurement and testing of designed circuits seems to be the next necessary stage in teaching circuit synthesis.

Real experimental laboratory requires corresponding equipments for given experiments and adequate time schedule to enable students performing the measurements. In praxis the main limitation can be especially convenient time schedule, because according security reasons the presence of students in laboratory require also presents a tutor. On the other hand more advantage is for students who prefer to work late in the evening or during weekend when university laboratories are usually closed.

The Project of measure stand wants to solve mentioned problems and attempt to implement three stage chain "theory – simulation – experiment" solving all here above mentioned problems and limitations to teaching subject "Basics of Electronics" at the Department of Electronics and Multimedia communications, Technical university of Kosice.

The practical realization of stages is:

1. Stage "Theory": students passing the theoretical lectures design basic circuits within numerical exercises.
2. Stage "Simulation": The designed "calculated" circuits are simulated using PSpice tools.
3. Stage "Experiment": The designed and simulated circuits are realized and measured. To save time and costs this stage is realized in remote access laboratory that is "open" 24 hour in a day and 7 days in a week.

III. HARDWARE ARCHITECTURE OF THE REMOTE ACCESS LABORATORY

The measuring stand contents real measuring instruments for simulate of real measure. Those instruments are being controlled locally by a stand/laboratory server. Instruments control commands and acquired data are collected/presented to user through dynamic web pages broadcasted by laboratory web server. A utilized web camera allows increasing the impression of real laboratory for user displaying image from front panels of instruments and view of whole measurement workplace. The instrumentation must be switch on 24 hour in a day for months what can lead especially to a damage of these instruments.

The alternative solution is utilization of multifunction data acquisition board (DAQ) installed directly in laboratory server. The board contain multiplexed analogue input channel with 16 inputs, that are used for acquiring analogue signal from circuit under test, 2 analogue outputs that can substitute two generators 2 -3 programmable counters that can be utilized as pulse generators or frequency meters and 8 – 16 programmable digital inputs/outputs that can control circuit under test by, e.g., controlling multiplexer changing component in the circuit, or they can be used for generation digital stimulus or acquiring digital signals from circuit under test.

Instrumentations based on DAQ cards require also software realizing data acquisition and presentation in the form of "virtual instrumentation". Virtual instruments are very simply integrated to dynamic web pages. Also web server administrating the web pages can be simply implemented in the same machine like performs the laboratory – measurement server role. In other words one PC can perform control of a few measurements stands using DAQ cards and virtual instruments and serves in parallel as web server. Required realization costs could be decreased at least 5 – 10 times in comparison with traditional solution utilizing stand-alone instruments.

Laboratory hardware consists of two DAQ boards (PCI 60336 and PCI 6251) with connector block SCC-68 where circuits under test are connected and FPGA circuits for testing of AD converter. In the present there are five such circuits:

1. Stabilized voltage source
2. Photovoltaic cell
3. Differential and Darlington transistor amplifiers
4. Operational amplifiers
5. Power amplifier
6. two AD converters

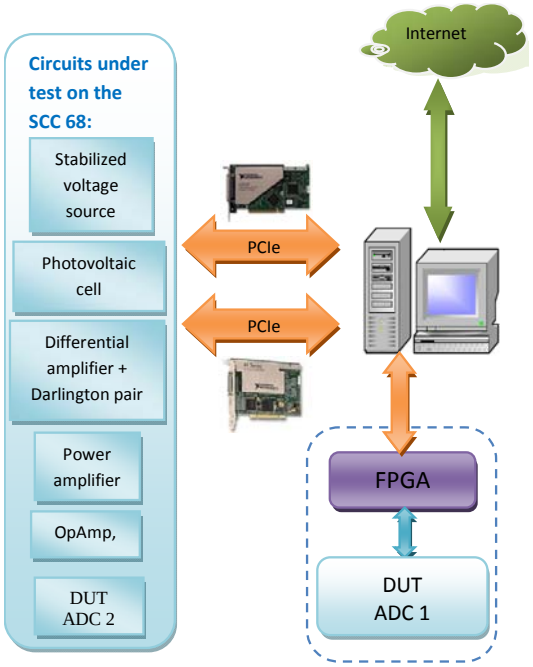


Fig. 1 Architecture of the proposed remote laboratory

The control PC is a common PC with Intel® Core™2 Quad processor, 4GB of memory and OS Windows XP. The PC performs both tasks – laboratory server as well as web server. The practical experiments showed that this simple solution powered by one simple low cost PC has high efficiency. When one student has control on the measuring stand processor works 10% of overall performance.

To increase the number of possible changes in component network of measured circuit, analogue multiplexers were implemented into the circuit. For example, the laboratory stand for testing operational amplifiers circuits consists of four independent circuits for four laboratory exercises. Each measuring circuits is being controlled by a multiplexer, which allow changing circuit components according the command by a student from remote computer. Control multiplexer as it is shown in Fig. 2 for inverting amplifier. The final value of feedback resistor is determined by parallel connection combination of four resistors connected through the analogue switching circuit.

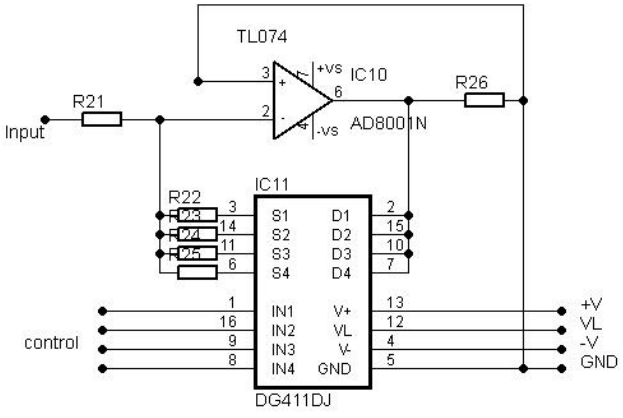


Fig. 2. Switching of feedback resistor in inverting circuits of operational amplifier

The latest working stand consist testing stand for testing basic parameters of AD converters. Testing of ADC converters take too much time and therefore it is need for reduce of this time. In these project problem was solved by use the FPGA circuits which do testing separately and gives to user only data which can be modified.

Workplace will use measuring card with analog and digital inputs and outputs. As AD converters will use a circuit AD574A. The AD574A is a complete 12-bit successive-approximation analog-to-digital converter with 3-state output buffer circuitry for direct interface to an 8- or 16-bit microprocessor bus. A high precision voltage reference and clock are included

on-chip, and the circuit guarantees full-rated performance without external circuitry or clock signals [5].

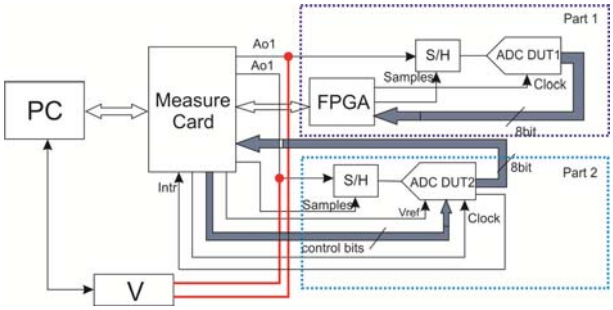


Fig. 3 Consist of measuring workplace for testing basic parameters of AD converters

In the first part is ADC connected to the card and set on default value of parameters. ADC tests both static and dynamic method. For this purpose stimulus signals has to be DC voltage and harmonic with adjustable amplitude. From the dynamical method of testing is used histogram and FFT method. For testing by these method is used FPGA circuit. The FPGA circuit create clocks signal, sampling signal and collect digital data from the ADC. These data are sent to the measuring card as final group of data. In case of static method FPGA do all algorithm of looking for right transition level. By using this circuit we reduce time of testing and power of PC processor. Data obtained from FPGA are saved in a PC and after that controlling programs work with them.

First part of testing string allows use FFT , histogram a statically method consecutively. User will have possibility to save the data and perform analyzes with them. There is no possibility of the unauthorized control FPGA circuits by the user. FPGA is programmed and operative alone. Acquired data are memorialized from FPGA circuit in PC and they are available for testing.

Second part of the working stand is devoted for training implementation of ADC in various modes. The utilized mode is being selected remotely through the DAQ card.

DAQ card generates signal for control of ADC for the study of ADC. These signals can be: sine, ramp and triangle. Attributes of stimulus signal will be possible to change by user.

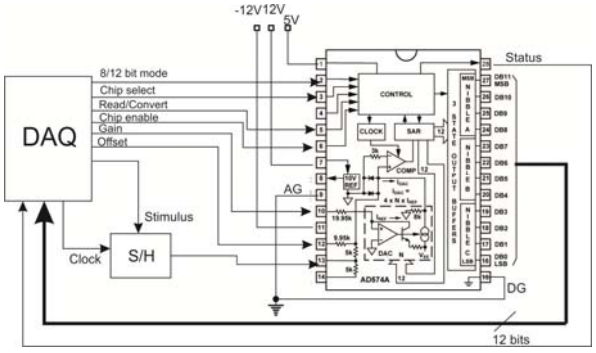


Fig. 4 Schematic second part of working stand

Figure 4 shown second part of testing stand. There will be AD converter in basic connection. The user will have possibility to study how various modalities of ADC could be settled. The possible nodes for selection are:

- input offset and gain
- switching of 8 or 12 bit operation mode
- selection for unipolar and bipolar mode of conversion.
- input analog voltage range
- chip select node (CS)
- read/convert node (RC)
- sampling frequency for sample and hold circuit

User can take fully control of AD converter by various input control bits. User uses contrloing program to control of this parts of working stand. The program obtain several preset modes of working ADC, also there is posible to create own working modes

IV. Control Software for Measurement Stands

The control software has been developed in LabVIEW by National instruments. LabVIEW was chosen because of its features: easy learning and effective programming, rich libraries supporting DAQ cards, signal processing and visualisation, etc. An additional advantage of LabVIEW is simple way of publish control program on the Internet. The only disadvantage of web server created in LabVIEW is that clients must have installed a special plug-in "Run Time Engine" that can be downloaded and installed to client web browser free [4].

LabVIEW web server also administrates the connection of every client. The first connected student asking for access to a measurement workplace gives the access automatically for unlimited time. Because we have five independent workplaces five students can work in parallel and independently. If another student connects to the server and asks for access to an occupied workplace, web server inserts this request to the queue of requests and limits the rest time for active user to a limit that could be set by administrator – usually 10-15 minutes. The active user is noticed that next user wants to take control of the measure stand and also is noticed that the workplace will be released in a time.

The design of front panels of VIs controlling particular workplaces was created in the form very similar to real instrumentation. User can control the virtual instruments as real ones and their behaviour is very similar to real traditional instruments.

A. Virtual generator

User interface of virtual signal generator is shown in Fig. 5



Fig. 5 User interface of virtual signal generator

B. Virtual digital multimeter

User interface of virtual digital multimeter is shown in Fig. 6.



Fig. 6 User interface of virtual digital multimeter

C. Control of other signals, selection of measuring nodes, presenting some other measured values

In workplace the changing component values, selection of measuring nodes and control and displaying some other non electrical values. All of this is realised by user interface in the form of special virtual instruments. The typical example is user interface developed for workplace with photovoltaic cell (Fig. 7.). The user controls here photovoltaic cell lighting, changes load resistance and measures light intensity by luxmeter realised by photoresistor.

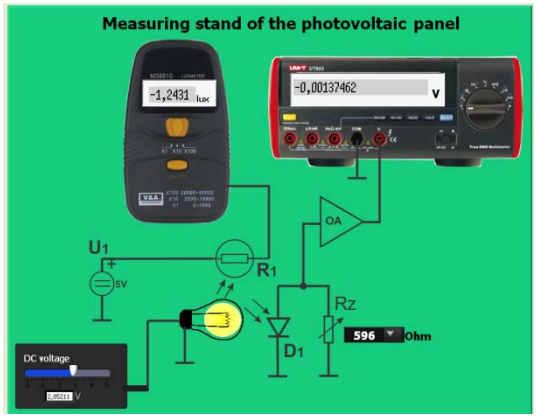


Fig.7 User interface of photovoltaic cell workplace

Another example is in Fig. 7. User can here set up the load of stabilised voltage source on the workplace.

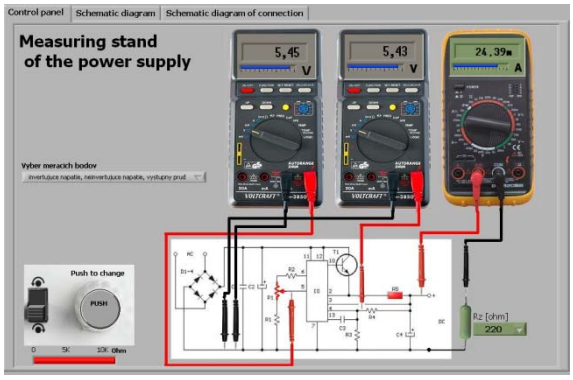


Fig. 8 User interface of power supply measure stand

V. CONCLUSIONS

Realization of remote controlled measurement laboratory for measuring basic electronic circuits as a support and the last stage of chain "theory – simulation – experiment" for effective teaching subject "Basics of Electronics" at the Department of Electronics and Multimedia communications, Technical university of Kosice is the process of testing, improvement and spreading in presence. In the summer semester of 2011 it was applied to regular teaching process at the department for the first time.

ACKNOWLEDGMENT

The work is a part of project supported by the Cultural and Educational Grant Agency of Slovak republic (029TUKE-4/2012)

This work was supported by Agency of the Ministry of Education of the Slovak Republic for the Structural Funds of the EU under the project Development of Centre of Information and Communication Technologies for Knowledge Systems (project number: 26220120030).

REFERENCES

[1] Daponte, P, Grimaldi, D., Rapuano S.: New Frontier of the m-Learning in Instrumentation and Measurement: Experiment Creation, Proc. Of 16th International Conference on Software, Telecommunications and Computer Networks, Split, CROATIA, SEP 25-27, 2008, pp. 412-416
[2] L. Perettoa, S. Rapuanob, M. Riccioc, D. Bonattia: Distance learning of electronic measurements by means of measurement set-up models, Measurement, Volume 41, Issue 3, April 2008, Pages 274-283
[3] D. Sampera, J. Santolaria, A. Majarenaa and J. Aguilar: Comprehensive simulation software for teaching camera calibration by a constructivist methodology, Measurement Volume 43, Issue 5, June 2010, Pages 618-630
[4] http://joule.ni.com/nidu/cds/view/p/id/2088/lang/en
[5] Analog Devices: datasheet "Complete 12-Bit A/D converter AD574A
[6] IEEE Std. 1241 – 2000, "IEEE Standard for Terminology and Tests Methods for Analog to digital Converters", IEEE, Inc. New Yourk, USA 2000
[7] Plassche,R.: Integrated Analog - to – Digital and Digital – to– Analog Converters. Kluwer Academic Publishers, 1994
[8] Razavi,B.:Principles of Data Conversion System Design, IEEE Press,1995

Steganalýza farebných statických obrazov na báze DWT transformácie

Branislav GALLIK, Gabriel BUGÁR, Vladimír BÁNOCI, Dušan LEVICKÝ

Katedra elektroniky a multimediálnych telekomunikácií, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

branislav.gallik@student.tuke.sk, gabriel.bugar@tuke.sk, vladimir.banoci@tuke.sk, dusan.levicky@tuke.sk

***Abstrakt* — Predložený článok sa zaoberá problematikou steganalýzy farebných statických obrazov vo formáte JPEG, pomocou voľne dostupných steganografických metód s využitím DWT transformácie . Obsahuje opis navrhnutého steganalýzneho modulu na báze DWT transformácie, jednotlivé fázy jeho činnosti a analýzu dosiahnutých výsledkov.**

***Kľúčové slová* — Steganalýza, DWT, JPEG, klasifikácia.**

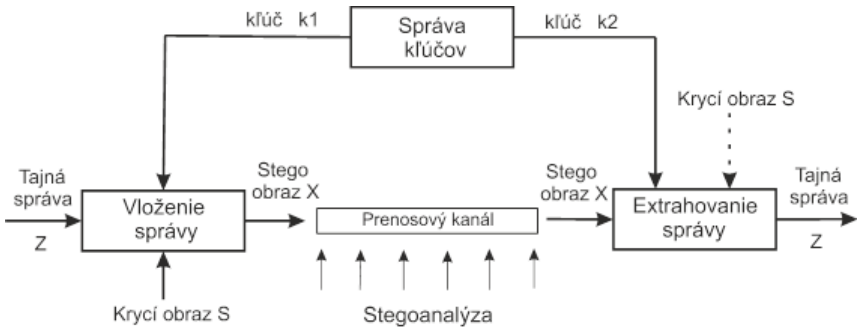
I. ÚVOD

Hlavným cieľom steganografie je ukryť tajnú správu do krycieho média, zatiaľ čo steganalýza vykonáva detekciu prítomnosti tajnej správy v krycom médiu[4]. Ako krycie médium sa veľmi často využívajú statické obrazy a tieto sa označujú sa ako stego obrazy. Cieľom steganalýzy je zisťovanie štatistických anomálií stego obrazov, ktoré sú založené na extrahovaní charakteristických vlastností z typických krycích obrazov bez modifikácie, ktorú spôsobí vloženie tajnej správy. Nemodifikované a neskreslené krycie obrazy obsahujú predvídateľnú štatistickú koreláciu a vykonaním rôznych úprav resp. modifikácií krycích obrazov dôjde k naručeniu tejto korelácie. Medzi najznámejšie spôsoby modifikácie obrazov patrí modifikácia globálneho histogramu, modifikácia vnútro blokovej a medzi blokovej závislosti alebo štatistikami prvého rádu a vyšších rádov. Väčšina algoritmov steganalýzy je založených na využívaní silných závislostí medzi obrazovými prvkami, ktoré sú typické pre prirodzené obrazy [1].

II. VŠEOBECNÝ MODEL OBRAZOVEJ STEGANOGRAFIE

Na Obr. 1 je znázornený všeobecný model obrazovej steganografie a steganalýzy. Odosielateľ vloží tajnú správu Z do zvolených oblastí krycieho obrazu S pomocou vkladacieho algoritmu s vkladacím kľúčom k1, ktorý vytvára stego obraz X. Stego obraz X je potom prenášaný verejným prenosovým kanálom k príjemcovi. Príjemca z prijatého stego obrazu X extrahuje tajnú správu Z pomocou algoritmu pre extrakciu s kľúčom extrakcie k2. Nepovolaná osoba , ktorá vykonáva stegoanalýzu, má k dispozícii iba podozrivé obrazy prenášané cez prenosové kanály, nepozná kľúče k1, k2, krycí obraz S a algoritmus vloženia tajnej správy do S. Vo všeobecnosti platí, že vloženie tajnej správy Z do S môže byť modelované ako aditívny šum, preto vzťah medzi stego obrazom X, krycím obrazom S a tajnou správou Z je daný vzťahom X=S+Z, kde Z je zvyčajne nezávislé od S.

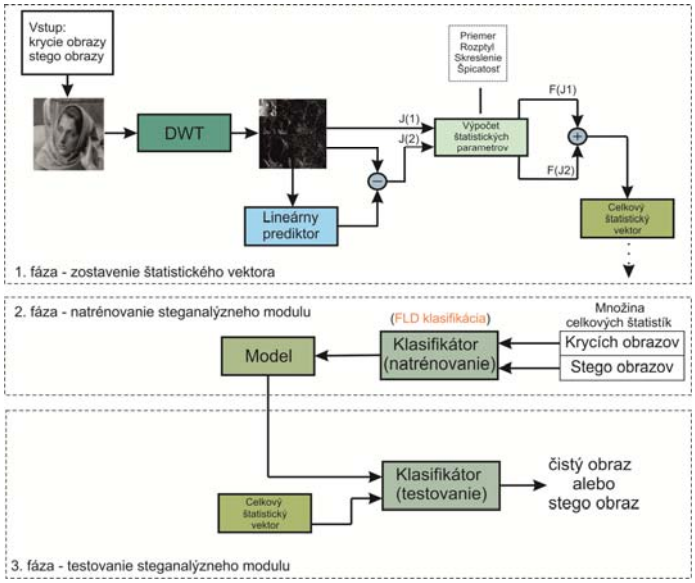
ISBN 978-80-553-0890-6 © 2012 FEI TUKE



Obr. 1 Všeobecný model obrazovej steganografie a steganalýzy

III. NÁVRHNUTÝ STEGANALÝZNY MODUL NA BÁZE DWT

Navrhnutý steganalýzny modul na báze DWT transformácie slúžiaci na detegovanie vložených tajných správ vo farebných obrazoch formátu JPEG [6]. Jeho činnosť pozostáva z troch hlavných fáz, a to z fázy zostavenia štatistického vektora, fázy trénovania a z fázy testovania. Na Obr. 2 je znázornená bloková schéma navrhnutého steganalýzneho modulu na báze DWT transformácie.



Obr. 2 Bloková schéma navrhnutého steganalýzneho modulu na báze DWT transformácie

A. Zostavenie štatistického vektora

Prvá fáza činnosti steganalýzneho modulu pozostáva z výpočtu štatistických parametrov krycích obrazov a stego obrazov a následného zostavenia štatistického vektora.

Obrazové prvky vstupného farebného obrazu, buď krycieho alebo stego obrazu, sa rozložia pomocou trojúrovňovej 2D DWT do oblasti transformačných koeficientov. Potom sa pre každú vertikálnu, horizontálnu, diagonálnu oblasť, každej úrovne rozkladu a pre každú zložku (RGB) vypočítajú štatistické parametre ako priemer, rozptyl, skreslenie a špicatosť, čím sa získa časť celkového štatistického vektora s dĺžkou 36(n-1)=36(4-1)=108, ktorý slúži na rozlíšenie krycích obrazov od stego obrazov.

Ďalším krokom je zistenie logaritmickej chyby medzi skutočnými transformačnými koeficientmi a predpokladanými koeficientmi, získanú pomocou navrhnutého lineárneho prediktora, ktorý reprezentuje kalibračný prvok steganalýzneho modulu. Pre každú takto získanú logaritmickú chybu, pre každú vertikálnu, horizontálnu, diagonálnu oblasť, každej úrovne rozkladu a pre každú zložku (RGB) sa vypočítajú štatistické parametre ako priemer, rozptyl, skreslenie a špicatosť. Tým sa vytvorí druhá časť celkového štatistického vektora s dĺžkou 36(n-1)=36(4-1)=108, ktorú tvorí súbor chybových štatistík.

Celková dĺžka štatistického vektora použitého pre rozlíšenie čistých obrazov od stego obrazová teda bude rovná 216.

B. Fáza natrénovania

Úlohou druhej fázy činnosti steganalýzneho modulu je vytvoriť model na základe natrénovania rozličných krycích obrazov a stego obrazov, ktorý sa neskôr použije v poslednej fáze.

Získané celkové štatistické vektory množín krycích obrazov a celkové štatistické vektory množín stego obrazov vstupujú do bloku klasifikátora, ktorého základnou časťou je neparametrická klasifikácia FLD. Podľa tejto klasifikácie sa vytvorí model natrénovaných krycích obrazov a stego obrazov.

C. Fáza testovania

Poslednou fázou činnosti navrhnutého steganalýzneho modulu je fáza testovania, ktorá rozhodne, či daný obraz obsahuje alebo neobsahuje vloženú tajnú správu, teda či bol obraz podrobený steganografickému algoritmu.

Z obrazu, ktorý je podrobený testovaniu, sa najprv získa celkový štatistický vektor podľa prvej fázy steganalýzneho modulu. Získaný celkový štatistický vektor spolu s natrénovaným modelom vstupuje do bloku klasifikátora, ktorého základom tvorí FLD klasifikácia [6]. Následne blok klasifikátora rozhodne o tom, či testovaný obraz patrí do skupiny stego obrazov alebo čistých obrazov.

IV. ZÁVER

Testované krycie obrazy formátu JPEG boli vytvorene pomocou niekoľkých druhov fotoaparátov. Podobne vznikla aj databáza označených a testovaných obrazov, pričom pri vytváraní označených obrazov bola snaha čo najrôznorodejšie meniť kapacitu vloženej tajnej správy v krycom obraze. Celková takto vytvorená databáza obsahovala približne 32 000 obrazov.

Hlavným cieľom bolo navrhnúť a vytvoriť steganalýzny modul na báze diskrétnej waveletovej transformácie, ktorého úlohou je detekovať tajné správy vložené pomocou steganografického algoritmu obsiahnuté v statickom farebnom obraze vo formáte JPEG. Objektom testovania boli testované množiny označených obrazov , ktoré boli vytvorené pomocou steganografických algoritmov ako F5, JPHS, MB1, MB2, MHF, MHF_dz, Outguess. Výsledkom hodnotenia kvality navrhnutého steganalýzneho modulu je percentuálny údaj hovoriaci o množstve obrazov danej testovanej množiny, ktoré boli modulom vyhodnotené ako označené obrazy.

Postup pri testovaní je taký, že najprv sa pomocou FLD klasifikácie natrénujú určité množstvá krycích obrazov s určitým druhom označených obrazov a testovanie sa vykonáva na rovnakom druhu označených obrazov alebo krycích obrazov testovaných množín s použitím FLD klasifikácie. Týmto spôsobom testovania je možné zistiť, či model označovaný ako cover_MHF, ktorý vznikol natrénovaním krycích obrazov s označenými obrazmi vzniknutými pomocou steganografického algoritmu MHF, dosahuje najväčšie hodnoty detekcie označených obrazov testovanej množiny, a to približne 52%. Najmenšie hodnoty detekcie označených obrazov testovanej množiny, približne 16%, dosahuje model označovaný ako cover_MB1, ktorý vznikol natrénovaním krycích obrazov s označenými obrazmi vytvorenými pomocou steganografického algoritmu MB1. Z výsledkov možno usúdiť, že najväčšie hodnoty detekcie krycích obrazov testovacej množiny, a to až 24%, vykazoval natrénovaný model označovaný ako cover_MB2, a najmenšie hodnoty detekcie krycích obrazov natrénovaný model cover_JPHS, približne 8%.

Pri vytvorenom modeli označovanom ako cover_all sa postupovalo tak, že najprv sa natrénujú pomocou FLD klasifikácie všetky množiny krycích obrazov so všetkými druhmi označených obrazov a testovaniebolo vykonávané na každom druhu označených obrazov a krycích obrazov testovaných množín s použitím FLD klasifikácie. Z experimentálnych výsledkov možno usúdiť, že najväčšie hodnoty detekcie označených obrazov testovanej množiny, a to až 89%, vykazovala množina označených obrazov pomocou algoritmu F5 a najmenšie hodnoty množina označených obrazov pomocou algoritmu JPHS, približne 5%. Hodnota detekcie krycích obrazov testovanej množiny natrénovaným modelom cover_all sa pohybovala okolo 1%.

Výhodou navrhnutého steganalýzneho modulu je jeho použiteľnosť na najrozšírenejší obrazový formát JPEG, nízka výpočtová náročnosť, rýchle natrénovanie modelu z už vopred vypočítaných štatistík. Nevýhodou modulu je použitá lineárna neparametrická klasifikácia FLD, nízka hodnota detekcie tajnej správy v krycom obraze, veľkosť databázy a množstvo obrazov, rovnako ako aj manipulácia s veľkým objemom dát a výsledkov.

POĎAKOVANIE

Táto práca bola vytvorená realizáciou projektu Rozvoj Centra informačných a komunikačných technológií pre znalostné systémy (kód ITMS projektu: 26220120030) na základe podpory operačného programu Výskum a vývoj financovaného z Európskeho fondu regionálneho rozvoja.

LITERATÚRA

[1] KUMAR, Mahendra: Steganography and steganalysis of JPEG images. Florida: University of Florida, Department of computer and information sciences and engineering, 2011.

[2] WESTFELD, Andreas: F5 – steganografický algoritmus. In IHW ’01: Proceedings of the 4th international workshop on information hiding, strany 289 –302. Springer-Verlag, 2001.

[3] [9] GUJAR, Sujit Prakash: A project report: Measures for classification and detection in steganalysis. Bangalore, India: Indian Institute of Science, Computer Science and Automation, 2006.

[4] MORKEL, T. – ELOFF, J.H.P. – OLIVIER, M.S. :An Overview of Image Steganography. Proceedings of the Fifth Annual Information Security South Africa Conference (ISSA2005).

[5] PROVOS, Niels: Outguess – steganografický algoritmus. <http://www.outguess.org/outguess-0.2.tar.gz>.

[6] Gallik, B.: Steganalýza v statických obrazov na baze DWT transformacie. Diplomová praca, KEMT, 2012

Digitálna vodotlač vo videu na báze hybridného algoritmu DCT-SVD

Peter GOČ-MATIS, Martin BRODA, Dušan LEVICKÝ

Katedra elektroniky a multimediálnych telekomunikácií, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

peter.goc-matis@tuke.sk, martin.broda@student.tuke.sk, dusan.levicky@tuke.sk

***Abstrakt* — Predkladaný článok sa zaoberá digitálnou vodotlačou vo videu v reálnom čase s využitím SVD a DCT transformácií. Navrhovaná metóda digitálnej vodotlače vo videu je založená na poznatkoch známych z digitálnej vodotlače v statických obrazoch. Predkladaný článok obsahuje krátky opis metód digitálnej vodotlače a tiež popisuje útoky na digitálne vodoznaky špecifické pre video. Článok taktiež popisuje proces vkladania a extrakcie digitálneho vodoznaku na báze hybridného algoritmu DCT- SVD. Posledná časť článku opisuje experimenty vykonávané na videách označených navrhnutou metódou digitálnej vodotlače. Cieľom experimentov bol test robustnosti navrhovanej metódy voči útokom na vodoznak špecifických pre video.**

***Kľúčové slová* — digitálna vodotlač, vodoznaky vo videu, transformovaná oblasť, blok extrakcie vodoznaku, SVD transformácia**

I. ÚVOD

Dnešná doba prináša so sebou masívne rozšírenie spracovania a prenosu multimediálnych dát, ako aj široký rozmach sieťových technológii. Analógová forma multimédií bola prakticky úplne nahradená digitálnou formou, a to takmer vo všetkých oblastiach ľudského života.

Digitálna forma dát umožňuje jednoduchý záznam, spracovanie a prenos multimediálnych dát cez komunikačné prostredie. Na druhej strane ale stoja nevýhody digitálnej formy dát, a to hlavne možnosť jednoduchého kopírovania týchto dát bez pozorovateľnej zmeny kvality vytvorenej kópie. Takto vytvorené kópie môžu byť vďaka vysokým prenosovým rýchlostiam súčasných sietí nelegálne distribuované prakticky do celého sveta. Možnosť tohto nelegálneho šírenia predstavuje veľkú nevýhodu pre autorov a distribútorov digitálnych multimedií, ktorí takto prichádzajú o svoje zisky [1].

II. DIGITÁLNA VODOTLAČ

Digitálna vodotlač predstavuje techniku vloženia prídavnej informácie do rôznych druhov multimédií. Dôležitou vlastnosťou vkladanej informácie je jej zmyslová nepostrehnuteľnosť a odolnosť voči potenciálnemu poškodeniu [2][3].

Prídavná informácia v sebe môže niesť údaje o vlastníkovi autorských práv, resp. o distribútorovi a taktiež slúži ako prostriedok na zabezpečenie integrity dát.

Vkladaný vodoznak môže mať rôznu formu a podobu v závislosti od použitia v aplikácii. Môže mať formu konečnej postupnosti symbolov, resp. číslic, prídavnej obrazovej informácie (loga), segmentu rečového signálu, poprípade môže mať formu 1-bitovej informácie, ktorá indikuje, či originálne dáta sú alebo nie sú označené vodoznakom.

Vodoznak vkladaný do multimediálnych dát musí spĺňať tri základné požiadavky, a to [4]:

- robustnosť,
- perceptívna transparentnosť,
- kapacita.

Pri návrhu metódy digitálnej vodotlače je nutné voliť vyhovujúci kompromis medzi týmito tromi požiadavkami.

A. Digitálna vodotlač vo videu

Metódy digitálnej vodotlače vo videu môžeme rozdeliť do troch základných skupín [5][6]:

- metódy na báze využitia digitálnej vodotlače v statických obrazoch,

- metódy na báze využitia časovej dimenzie vide,
- metódy na báze využitia formátu komprimovaného videa.

B. Útoky na digitálne vodoznaky vo videu

Útoky na digitálnu vodotlač vo videu predstavujú všetky úmyselné aj neúmyselné operácie, ktoré sú vykonávané útočníkom s cieľom odstrániť vodoznak z označených dát a získať tak prístup k neoznačenému multimediálnemu obsahu. Medzi špecifické útoky, na digitálne vodoznaky vo videu patria:

- vynechanie snímok,
- výmena snímok,
- spriemerňovanie snímok,
- útoky na báze štatistickej analýzy.

Medzi útoky na vodoznaky špecifické pre video patria aj filtrácia videa pre rôzne korekcie obrazu, gama korekcia na zlepšenie kvality obrazu a taktiež transformáciu videa z jedného komprimovaného formátu na iný. Primárnym cieľom týchto operácii nie je zničiť alebo poškodiť vodoznak, a preto sa označujú ako neúmyselné útoky.

III. NAVRHNUTÁ METÓDA DIGITÁLNEJ VODOTLAČE

A. SVD transformácia

Singulárna dekompozícia (SVD) patrí medzi novšie metódy a využíva sa pri matematických výpočtoch s maticami alebo pri prehľadávaní vo veľkom množstve dát.

SVD transformácia je ortogonalizačný proces, ktorý vedie ku dekompozícii matice *A* na diagonálnu maticu *S* a jej ľavé a pravé singulárne matice *U* a *V,* pričom platí [7][8]

$$A = U.S.V^T \tag{1}$$

Za predpokladu, že matica *A* je rozmeru mxn môžno vzťah zapísať v tvare

$$SVD\left(A_{m\times n}\right) = \left[U_{m\times n} S_{m\times n} V_{n\times n}\right] \tag{2}$$

Vynásobením *U*,*S* a *V* (kde V je transponovaná matica) sa získa matica

$$A_{m\times n} = U_{m\times n}.S_{m\times m}.V^T_{n\times n} = \sum_{i=1}^{\min(m,n)} \sigma_i.u_i.v_i^T \tag{3}$$

Kde $\sigma_i = \{R+\}$, *i=1 ... min(m,n)* sú singulárne hodnoty, t.j. diagonálne prvky matice *S*, ktoré sú zoradené v zostupnom poradí. u_i sú ľavé singulárne vektory, t.j stĺpce matice *U* a v_i sú pravé singulárne vektory, t.j. riadky matice V_T. *U* a *V* sú unitárne matice, čo znamená že platí

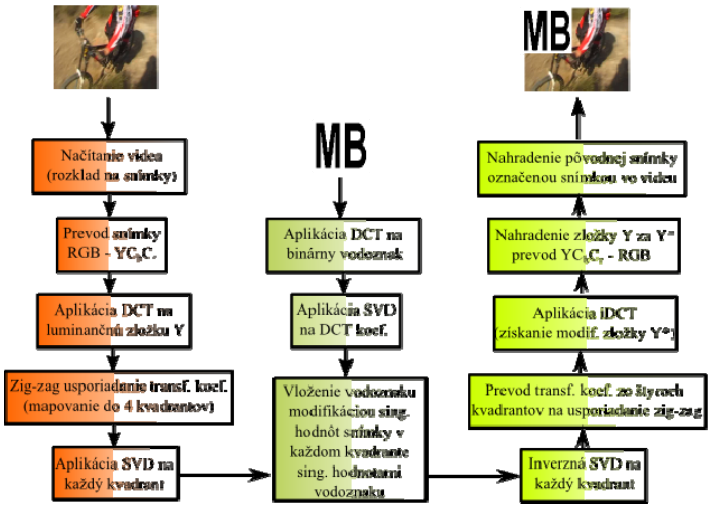
$$U.U^T = I_{m\times m}, V.V^T = I_{m\times n} \tag{4}$$

kde $I_{m\times m}$ a $I_{m\times n}$ sú jednotkové matice, V^T je transponovaná matica V .

Zvýšenie signulárnych hodnôt *(SV)* matice *S* vedie k zvýšeniu jasu obrazu. Hodnoty matice *S* sú úzko späté s s jasom obrazu, zatiaľ čo skutočná geometria obrazu závisí na unitárnych maticiach *U* a *V*, ktoré predstavujú horizontálne a vertikálne okraje obrazu. Jednou z veľkých výhod použitia SVD je invariantnosť singulárnych hodnôt voči bežným operáciám spracovania obrazu ako aj odolnosť voči geometrickým zmenám, napr. rotácia a taktiež odolnosť voči škálovaniu.

B. Blok vkladania vodoznaku

Blok vkladania vodoznaku vykonáva načítanie videa, SVD a DCT transformáciu, vkladanie vodoznaku a inverznú DCT a SVD transformáciu. Vstupmi pre tento blok sú originálne video a vkladaný vodoznak. Blok vkladania vodoznaku je znázornený na Obr. 1.



Obr. 1 Bloková schéma vkladania vodoznaku

V experimentoch boli použité 3 videosekvencie s rozdielnou dynamikou scén (dynamické, stredne dynamické a málo dynamické). Rozlíšenie videosekvencií bolo 352x288 obrazových prvkov. Videosekvencie boli v nekomprimovanom formáte v „true color“ móde so snímkovou frekvenciou 25 snímok za sekundu.

Vkladaný vodoznak je binárny obraz s iniciálami M.B., ktorý je znázornený Obr. 2. Veľkosť vodoznaku je nutné modifikovať v závislosti na rozlíšení označovaného videa (v našom prípade 176 x 144 op).



Obr. 2 Vkladaný vodoznak

Video snímky sú konvertované z farebného modelu RGB do YcbCr. Pre modifikáciu vodoznakom bola použitá iba jasová zložka Y. Modifikácia farbonosných zložiek by totiž viedla k strate informácie pri podvzorkovaní týchto zložiek. Vodoznak je vkladaný do vopred vybraných snímok kvôli zníženiu vypočtovej a časovej náročnosti.

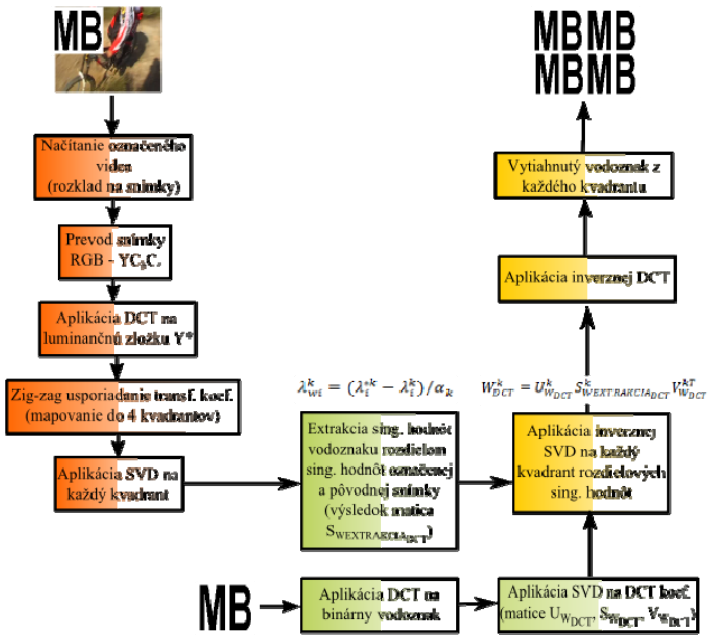
Proces vloženia vodoznaku môže byť popísaný nasledujúcim vzťahom

$$\lambda_i^{*k} = \lambda_i^k + \alpha_k \lambda_{wi}^k \qquad i=1,2,\ldots,n \tag{5}$$

kde koeficienty λ_i^{*k} sú modifikované singulárne hodnoty matice S^*, λ_i^k predstavujú pôvodné singulárne hodnoty snímky v danom kvadrante, λ_{wi}^k sú singulárne hodnoty matice S_{WDCT} daného vodoznaku a α_k sú škálovacie faktory, ktoré ovplyvňujú silu modifikácie singulárnych hodnôt, a teda voľba vyššej hodnoty spôsobuje zvýšenú odolnosť a robustnosť vodoznaku ale na druhej strane znižuje jeho perceptívnu transparentnosť.

C. Blok extrakcie vodoznaku

Blok extrakcie vodoznaku vykonáva extrakciu vodoznaku zo všetkých označených snímok, a to za účelom zvýšenia odolnosti voči útokom ako sú výmena a vynechanie snímok. Vstupmi pre tento blok sú originálne video, označené video a vkladaný vodoznak. Blok extrakcie vodoznaku je znázornený na Obr. 3.



Obr. 3 Bloková schéma extrakcie vodoznaku

Proces extrakcie vodoznaku môže byť popísaný nasledujúcim vzťahom

$$\lambda^{k}_{\text{wextrakcia}} = \left(\lambda^{*k}_{i} - \lambda^{k}_{i}\right)/\alpha_k \tag{6}$$

pre i=1, 2, ..., n, kde λ^{*k}_{i} sú singulárne hodnoty matice S^{*k}_{α}, ktorá bola vyjadrená z označenej snímky videa, λ^{k}_{i} predstavuje singulárne hodnoty matice S^{k}_{α} pre snímku z originálneho (neoznačeného) videa a α_k sú škálovacie faktory vyjadrujúce silu modifikácie pri vkladaní vodoznaku.

IV. Experimentálne výsledky

Navrhnutá metóda digitálnej vodotlače bola testovaná proti úmyselným a neúmyselným útokom. Kvalita extrahovaných vodoznakov po útokoch bola posudzovaná na základe objektívnych mier hodnotenia kvality, ktorými boli stredná kvadratická chyba (MSE), špičkový pomer signál/šum (PSNR) a taktiež bitová zhoda (BZ).

Vplyv procesu vkladania na označené dáta bol meraný pomocou PSNR a MSE, pričom hodnoty sú uvedené v Tabuľka 1.

Škálovací faktor α je úzko spojený s robustnosťou vloženého vodoznaku. Zvýšenie jeho hodnoty vedie k zníženiu hodnoty PSNR.

Tabuľka 1
Porovnanie PSNR a MSE pre rôzne typy videí z hľadiska dynamiky a pre rôzne škálovacie faktory

škálovacie faktory	**Testované video**			
		dynamické	**stredne dynamické**	**málo dynamické**
$\alpha_1 = 2$ $\alpha_{2\text{-}4} = 0.8$	**PSNR [dB]**	52,2274	51,7344	52,4301
	MSE	0,3894	0,4362	0,3739
$\alpha_1 = 5$ $\alpha_{2\text{-}4} = 1.5$	**PSNR [dB]**	47,2123	46,5276	47,414
	MSE	1,2355	1,4465	1,1794
$\alpha_1 = 10$ $\alpha_{2\text{-}4} = 3$	**PSNR [dB]**	41,662	40,8799	41,7765
	MSE	4,4306	5,31	4,3195
$\alpha_1 = 20$ $\alpha_{2\text{-}4} = 4$	**PSNR [dB]**	37,8417	37,0232	37,9699
	MSE	10,6884	12,9051	10,3773

Prvým útokom vykonávaným na označenom videu bolo spriemernenie snímok. Výsledky

bitovej zhody po extrakcii vodoznaku po vykonaní tohto útoku sú znázornené v Tabuľka 2.

Videá boli označené vodoznakom pri škálovacom faktore $\alpha_1=5$, $\alpha_{2\text{-}4}=1.5$. V experimente sa realizoval priemer 2, 3, 5 a 10 snímok z rovnakej scény videa. Ako je možné vidieť z tabuľky metóda je odolná voči tomuto typu útoku.

Tabuľka 2
Porovnanie kvality extrahovaného vodoznaku po útoku spriemernenia snímok

Typ videa podľa dynamiky	**Číslo kvadrantu**	**Počet spriemernených snímok**			
		2 snímky	3 snímky	5 snímok	10 snímok
		BZ [%]	**BZ [%]**	**BZ [%]**	**BZ [%]**
dynamické	1.	89,03	87,87	86,08	81,43
	2.	88,45	86,41	84,64	83,5
	3.	90,59	87,71	83,94	79,66
	4.	97,64	95,38	90,59	83,21
stredne dynamické	1.	87,2	82,73	80,17	79,7
	2.	96,3	94,26	92,91	89,43
	3.	99,08	97,71	95,27	90,76
	4.	99,73	98,66	95,27	86,27
málo dynamické	1.	86,25	86,01	85,68	85,48
	2.	83,9	82,88	82,07	81,57
	3.	85,9	83,83	82,05	80,31
	4.	94,12	92,05	89,58	85,77

Výsledky bitovej zhody po útoku sú veľmi dobré dokonca aj pri spriemernení 10 snímok. Spriemernenie takéhoto počtu snímok ale vedie k značenej degradácii kvality videa.

Ďalším z vykonávaných útokov bol útok kompresie. Ako kompresné štandardy boli vybrané H.264/MPEG-4 AVC a VP8 s bitovým tokom 1500 kbps.

Výsledky bitovej zhody po extrakcii vodoznaku po vykonaní útokov kompresie sú znázornené v Tabuľka 3 a Tabuľka 4.

Tabuľka 3
Porovnanie kvality extrahovaného vodoznaku po útoku kompresiou H264

Typ videa podľa dynamiky	**Číslo kvadrantu (extrahovaný vodoznak)**	**Škálovacie faktory**		
		$\alpha_1=2$ $\alpha_{2\text{-}4}=0.8$	$\alpha_1=5$ $\alpha_{2\text{-}4}=1.5$	$\alpha_1=10$ $\alpha_{2\text{-}4}=3$
		BZ [%]	**BZ [%]**	**BZ [%]**
dynamické	1.	99,97	100	100
	2.	99,74	99,84	99,96
	3.	99,97	100	100
	4.	80,88	86,77	96,66
stredne dynamické	1.	100	100	100
	2.	100	100	100
	3.	100	100	100
	4.	97,88	99,89	100
málo dynamické	1.	100	100	100
	2.	100	100	100
	3.	100	100	100
	4.	100	100	100

Tabuľka 4
Porovnanie kvality extrahovaného vodoznaku po útoku kompresiou VP8

Typ videa	**Číslo**	**Škálovacie faktory**

podľa dynamiky	**kvadrantu (extrahovaný vodoznak)**	α_1=2 $\alpha_{2\text{-}4}$=0.8	α_1=5 $\alpha_{2\text{-}4}$=1.5	α_1=10 $\alpha_{2\text{-}4}$=3
		BZ [%]	**BZ [%]**	**BZ [%]**
dynamické	1.	93,79	99,87	100
	2.	85,16	100	99,92
	3.	90,59	93,65	96,65
	4.	75,87	81,38	85,36
stredne dynamické	1.	99,86	100	100
	2.	99,98	100	100
	3.	99,9	99,95	99,87
	4.	90,12	96,74	98,84
málo dynamické	1.	99,99	100	100
	2.	100	100	100
	3.	99,71	100	100
	4.	95,65	99,17	99,45

Posledným vykonávaným útokom bol geometrický útok rotácie. Na videá bola aplikovaná rotácia o 1, 3 a 5 stupňov. Výsledky bitovej zhody po vykonaní tohto útoku su znázornené v Tabuľka 5.

Je možné sledovať, že so zvyšovaním rotácie, klesá bitová zhoda extrahovaného vodoznaku s originálom, avšak do hodnoty 5 stupňov bolo ešte možné vodoznak subjektívne rozoznať aspoň v jednom kvadrante. Rotácie o väčší počet stupňov nemajú v praxi žiaden zmysel, keďže potom už je daná zmena rozoznateľná voľným okom, a teda hneď je možné povedať, že na daný obraz bol realizovaný útok rotácie.

Tabuľka 5
Porovnanie kvality extrahovaného vodoznaku po geometrickom útoku rotácie

Typ videa podľa dynamiky	**Číslo kvadrantu (extrahovaný vodoznak)**	**Rotácia**		
		1 stupeň	3 stupne	5 stupňov
		BZ [%]	**BZ [%]**	**BZ [%]**
dynamické	1.	98,52	74,92	59,24
	2.	82,28	72,97	71,96
	3.	85,54	79,1	77,19
	4.	84,72	78,2	77,44
stredne dynamické	1.	96,23	83,86	71,63
	2.	96,12	81,56	76,14
	3.	98,44	92,17	83,5
	4.	97,75	86,31	81,8
málo dynamické	1.	99,42	84,68	69,94
	2.	85,83	72,59	69,23
	3.	85,98	79,28	74,44
	4.	89,77	77,02	75,24

V. ZÁVER

Experimentálne výsledky ukazujú, že navrhnutá metóda digitálnej vodotlače na báze DCT-SVD transformácie je robustná voči neumýselnému útoku kompresie vybranými kompresnými štandardmi.

Metóda je taktiež odolná voči vybraným úmyselným útokom, a to spriemernenia snímok a voči geometrickému útoku rotácie jednotlivých snímok.

Robustnosť voči jednotlivým útokom je možné zvýšiť modifikáciou škálovacieho faktora α, ktorý je s ňou priamo spojený. Treba však brať do úvahy fakt, že zvýšenie robustnosti vedie v zníženiu perceptívnej transparentnosti vloženého vodoznaku. Je teda nutné zvoliť vhodný kompromis medzi robustnosťou a perceptívnou transparentnosťou, a to v závislosti od aplikácie digitálnej vodotlače.

Hlavnou nevýhodou navrhnutej metódy je potreba originálnych dát v procese extrakcie vodoznaku.

POĎAKOVANIE

Táto práca bola vytvorená realizáciou projektu Rozvoj Centra informačných a komunikačných technológií pre znalostné systémy (kód ITMS projektu: 26220120030) na základe podpory operačného programu Výskum a vývoj financovaného z Európskeho fondu regionálneho rozvoja.

LITERATÚRA

[1] P. Goč – Matis, T. Kanócz, R. Ridzoň, D. Levický, "DWT in video watermarking.". In Research in Telecommunication Technologies 2010: the 12th International Conference: September 8-10, 2010, Velké Losiny, Czech Republic. - Ostrava : VŠB-TU, 2010 P. 219-222. - ISBN 978-80-248-2261-7.
[2] S. Katzenbeiser, F. Petitcolas, "Information Hiding techniques for Steganography and Digital Watermarking." Artech House, Boston, 2000.
[3] M. Miller, I. Cox, J. P. Linnartz, T. Kalker, "A review of watermarking principles and practices," In Digital Signal Processing in Multimedia Systems, chapter 17, pp. 461-485, 1999.
[4] M. Wu, B. Liu, Multimedia data hiding, Springer-verlag New York, Inc. 2003.
[5] T. Kanócz, T. Tokár, D. Levický, "Robust frame by frame video watermarking resistant against collusion attacks" In: Radioelektronika 2009. - Brno : Brno University of Technology, 2009 S. 99-102. - ISBN 9781424435364.
[6] G. Doerr, "Security Issue and Collusion Attacks in Video Watermarking." PhD. thesis, Universite de Nice Sophia-Antipolis, 2005.
[7] A. Basso, F. Bergadano, D. Cavagnino, "A Novel Block- based Watermarking Scheme Using the SVD Transform," January 2009.
[8] L. Rajab, T. Al-Khatib, A. Al-Haj, "Video watermarking Algorithms Using the SVD Transform," 2009.
[9] Broda,M.: Digitálna vodotlač s využitím SVD transformácie. Diplomová práca, KEMT, 2012

Typy a aplikácie optických korelátorov

Tomáš HARASTHY, Ľuboš OVSENÍK, Ján TURÁN

Katedra elektroniky a multimediálnych telekomunikácií, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

tomas.harasthy@tuke.sk, lubos.ovsenik@tuke.sk, jan.turan@tuke.sk

***Abstrakt* — Optické spracovanie dát patrí k najlepším možným riešeniam spracovania informácií, resp. obrazov s vysokým rozlíšením. Keď sa optika, ktorá využíva najrýchlejšie médium (svetlo) využije na spracovanie rovnakej informácie ako klasické elektronické systémy, čas potrebný na spracovanie rovnakého množstva informácie sa podstatne skráti. Najlepšie výsledky je však možno dosiahnuť spoluprácou týchto dvoch metód spracovania informácie. Príkladom takejto spolupráce je Optický korelátor, ktorý využíva rýchlosť svetla na transformáciu dvojrozmerných obrazov s vysokým rozlíšením. Optický korelátor je zariadenie, ktoré dokáže porovnávať dva obrazy vo veľmi vysokej rýchlosti (rýchlosť svetla). V tomto príspevku sú opísané druhy optických korelátorov s jednoduchými experimentmi. Optické korelátory sa vďaka ich vysokej rýchlosti spracovania informácie a pokrokom vo vývoji displejov na báze tekutých kryštálov stávajú čoraz viac komerčne dostupnejšie a žiadanejšie.**

***Kľúčové slová* — Fourierova transformácia, optická korelácia, prispôsobený filter, spojená transformácia**

I. Úvod

Optický korelátor (OC – Optical Correlator) je zariadenie, ktoré dokáže porovnávať dva obrazy alebo dvojrozmerne polia dát vo veľmi krátkom čase. Optická korelácia je definovaná ako výsledok Fourierovej transformácie dvoch funkcií *s(x,y)* a *r(x,y),* ktoré boli predtým Fourierovo transformované [1]. Existujú dva základné typy OC, prvý je založený na prispôsobenom filtri nazvaný Korelátor s prispôsobeným filtrom (MFC - Matched Filter Correlator) a druhý založený na spojenom výkonovom spektre - Korelátor so spojenou transformáciou (JTC - Joint Transform Correlator) [2,4]. Optická Fourierova transformácia (OFT – Optical Fourier Tranform) je jedným zo základných stavebných prvkov frekvenčnej analýzy obrazu. Táto analýza rozkladá obrazy podľa priestorových frekvencií čo umožňuje tiež jednoduchú frekvenčnú filtráciu obrazov [1]. Využitie OFT v optických korelátoroch bude podrobnejšie popísané v Kapitole 2. V Kapitole 3 budú porovnané MFC a JTC korelátory. V tejto kapitole budú tiež podrobne opísané princípy fungovania týchto korelátorov, uvedené ich výhody a nevýhody. Kapitola 4 sa bude venovať realizovaným korelátorom a ich aplikáciám. Záverečné zhrnutie článku je uvedené v Kapitole 5.

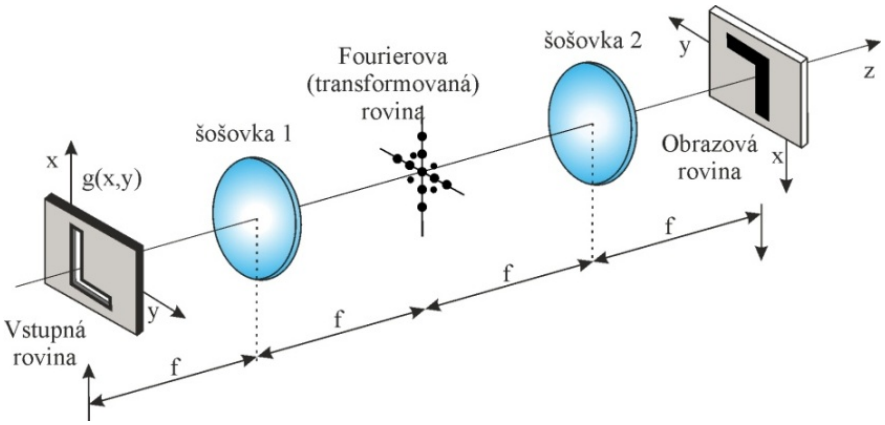
II. Optická Fourierova transformácia

A. Ukážka optickej Fourierovej transformácie

Dvojrozmerný signál (teda aj obraz) je možné vyjadriť ako sumu postupností harmonických funkcií [1]. V prípade obrazu sú tieto harmonické funkcie reprezentované striedaním jasu v obraze. Tento sínusový vzor môže byť kódovaný pomocou priestorovej frekvencie (frekvencia, ktorou je modulovaný jas v obraze), amplitúdy (rozdielu medzi najtmavšími a najjasnejšími špičkami v obraze, pričom tieto hodnoty môžu byť kladné aj záporné) a fázy (relatívny posun vĺn vzhľadom k pôvodnej sínusovej zložke). Keďže vo Fourierovom transformovanom obraze je uložená tá istá informácia ako vo vstupnom obraze, využitím spätnej FT môžeme získať úplne presnú repliku vstupného obrazu [1,2].

Ďalej vieme, že šošovka vytvára FT predmetu vždy vo svojej ohniskovej rovine, nezávisle od polohy predmetu vzhľadom na šošovku, na rozdiel od geometrického obrazu predmetu, ktorého poloha závisí od rovnice šošovky. Obr. 1 zobrazuje optickú Fourierovu transformáciu (OFT) pomocou spojných šošoviek v tzv. 4-f systéme (4-f znamená, že medzi vstupným a výstupným obrazom je vzdialenosť rovná štvornásobku ohniskovej vzdialenosti šošovky). Použitím druhej šošovky dostaneme inverznú OFT a tak tískame pôvodný obraz. Použitím maskovacieho filtra v transformovanej rovine dokážeme frekvenčne filtrovať daný obraz a zmeniť tak výstup 4-f

systému. Optické spracovanie (FT a filtrácia) je úplne paralelný proces výpočtu, ktorý využíva rýchlosť svetla. Keďže je proces paralelný jeho čas nenarastá so zvyšovaním rozlíšenia vstupného obrazu, ako to je pri elektronických výpočtových metódach (napr. FFT).



Obr. 1 Optická Fourierova transformácia v 4-f systéme.

III. OPTICKÉ KORELÁTORY

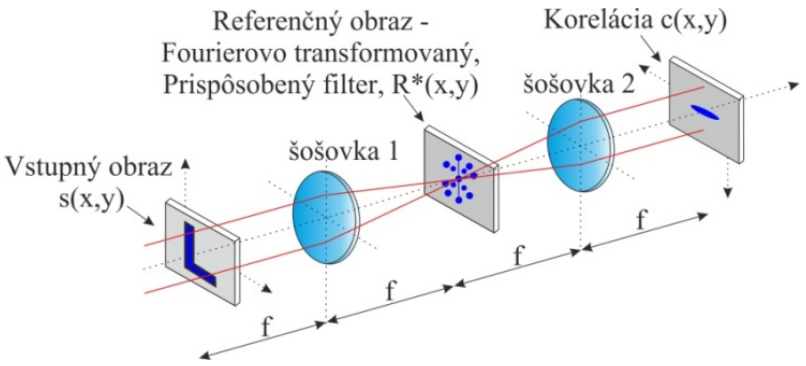
Existujú dva základné typy optických korelátorov (MFC a JTC). Korelátor s prispôsobeným filtrom (MFC) je najčastejšie využívaný. Jeho základ spočíva vo využití Vanderlugtovho filtra [1,3,4]. Základ Korelátora so spojenou transformáciou (JTC) spočíva vo využití Spojenej transformácie (Joint transform - JT) [1,3,4].

A. Korelátor s prispôsobeným filtrom (MFC)

Architektúra MFC vychádza z architektúry dvoj šošovkového systému opísaného vyššie. Je to najviac rozšírený typ korelátora, no zároveň je aj veľmi citlivý na presnú konštrukciu kvôli dodržaniu presných ohniskových vzdialeností šošoviek. Na Obr. 2 je znázornený princíp tohto korelátora. Korelácia zahŕňa dve nezávislé FT a to transformáciu vstupnej aj referenčnej zložky. Vstupná scéna je zobrazená na SLM a v obrazovej rovine šošovky 1 je vykonaná jej FT. V tej istej rovine je vložený aj komplexne združený FT obraz referenčnej scény, teda prispôsobený referenčný filter. Tento obraz je násobený transformovaným obrazom vstupnej scény. FT referenčnej scény je vykonávaná zvyčajne elektronicky a do optického systému už vstupuje ako filter. Šošovka 2 produkuje druhú FT súčinu transformovaných obrazov, výsledkom čoho je korelácia vstupnej a referenčnej snímky. Výhodou tejto metódy je vysoká priestorová šírka pásma produktu a extrémne rýchla realizácia. Hlavnou nevýhodou je elektronická realizácia prispôsobeného filtra z referenčnej scény .

Koreláciu môžeme zapísať v tvare vzťahu:

$$s \otimes r = c(x,y) = F^{-1}\{S(\alpha,\beta)H(\alpha,\beta)\} = F\{S(\alpha,\beta)R^*(\alpha,\beta)\}. \quad (1)$$



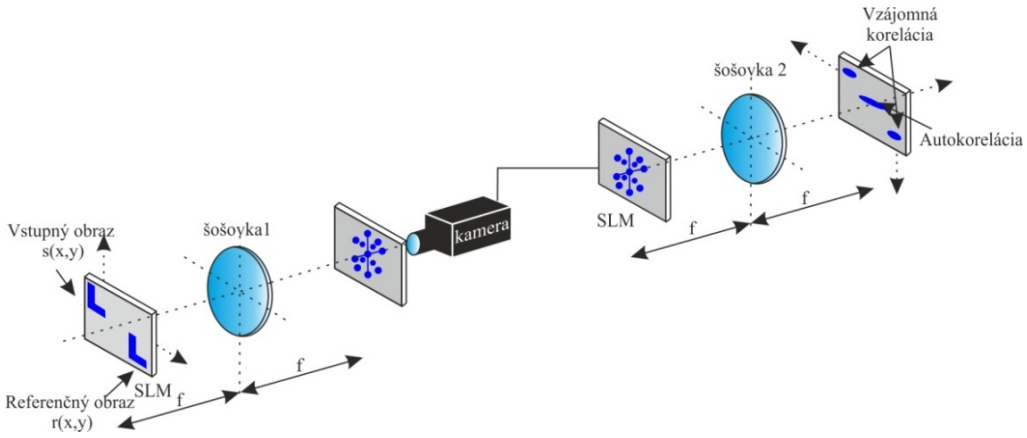
Obr.2 Princíp korelátora s prispôsobeným filtrom.

Operácia $\otimes$ predstavuje koreláciu a F^{-1} inverznú FT. Tu platí teoréma, že konvolúcia v časovej oblasti predstavuje koreláciu dvoch signálov (obrazov) vo frekvenčnej oblasti. Takto je možné pomocou FT získať koreláciu dvoch signálov [1,3].

B. Korelátor so spojenou transformáciou (JTC)

JTC sa líši od MFC najmä spôsobom usporiadania vstupnej a referenčnej zložky. Pri JTC sa

referenčná scéna netransformuje, ale je usporiadaná spolu so vstupnou scénou vedľa seba a zobrazená na vstupnom SLM. Následne sú obidva obrazy transformované spolu a produkujú tak Spojené výkonové spektrum (Joint Power Spectrum – JPS) (Obr. 3) [1,2,5]. Toto spektrum je potom prahované a znova transformované FT. Výstupom druhej korelácie je korelačná rovina, ktorá obsahuje korelačné špičky (ak bola na vstupe detegovaná zhoda). Špičky v korelačnej rovine určujú zhodu a dávajú nám dva druhy informácie o zhode vo vstupnom obraze. Intenzita korelačných špičiek nám hovorí o tom, aká silná je detegovaná zhoda a umiestnenie korelačných špičiek nám hovorí o tom ako sú tieto zhodné obrazy umiestnené vo vstupnej obrazovej rovine [2,5].



Obr.3 Princíp korelácie so spojenou transformáciou.

Vstupná rovina korelačného procesu môže byť zapísaná nasledovne:

$$v(x,z) = s(x,y) + r(x,y) \tag{2}$$

kde *s(x,y)* a *r(x,y)* sú vstupné a referenčné obrazy zobrazené vo vstupnej rovine a vytvárajú spojenú vstupnú scénu *v(x,y)*. Tento vstup je následne FT a vytvára tak *V(α,β)*.

$$V(\alpha,\beta) = F\{s(x,y)\} + F\{r(x,y)\} = S(\alpha,\beta) + R(\alpha,\beta). \tag{3}$$

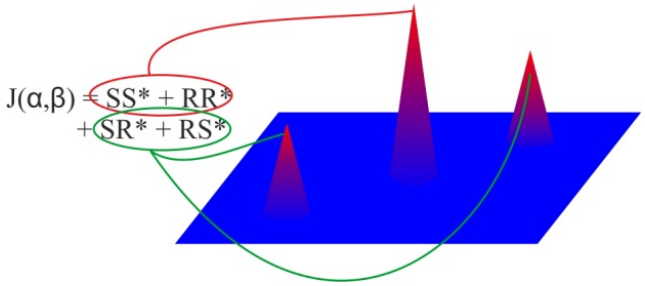
Ďalej vieme, že konvolúcia dvoch signálov je ekvivalentná násobeniu ich spektier. Súčin pri JTC predstavuje umocnenie absolútnej hodnoty komplexne sčítaných obrazov, ktoré produkuje spojené výkonové spektrum.

$$\begin{aligned} J(\alpha,\beta) &= |V(\alpha,\beta)|^2 = |S(\alpha,\beta) + R(\alpha,\beta)|^2 \\ &= |S(\alpha,\beta)|^2 + |R(\alpha,\beta)|^2 + 2|S(\alpha,\beta)R(\alpha,\beta)| \end{aligned} \tag{4}$$

Taktiež platí, že umocnenie absolútnej hodnoty komplexnej funkcie je rovné súčinu funkcie s jej komplexne združenou funkciou a teda platí:

$$\begin{aligned} J(\alpha,\beta) &= S(\alpha,\beta)\times S^*(\alpha,\beta) + R(\alpha,\beta)\times R^*(\alpha,\beta) \\ &+ S(\alpha,\beta)\times R^*(\alpha,\beta) + S^*(\alpha,\beta)\times R(\alpha,\beta). \end{aligned} \tag{5}$$

Výsledok rovnice 5 je znázornený na Obr. 4.



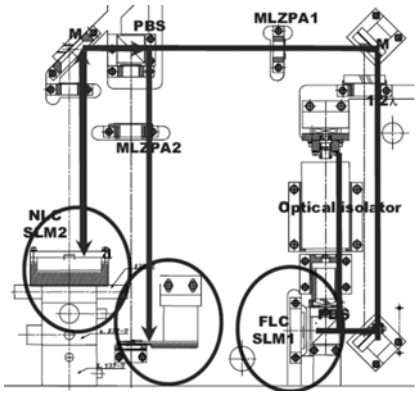
Obr.4 Korelačné špičky pri JTC korelátoroch.

IV. PRAKTICKÉ VYUŽITIE OPTICKÝCH KORELÁTOROV

MFC a JTC majú svoje výhody a nevýhody a podľa toho sú aj využívané v mnohých aplikáciách [1-5]. OC dokážu rozpoznávať tváre, veci, môžu využívať Fourierovu resp. Hilbertovu transformáciu a môžeme ich rozdeliť na dvoj resp. trojrozmerné korelátory. Podľa týchto kritérií sa využívajú v rôznych aplikáciách ako biometrické rozpoznávanie, rozpoznávanie dopravných značiek, sledovanie objektov a mnohé iné. Využitie OC je veľmi nápomocné najmä v aplikáciách, kde je potrebné rozpoznávanie v reálnom čase.

A. FARCO

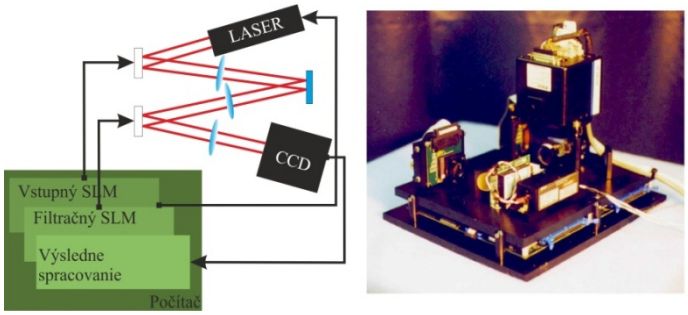
Fast Face-Recognition Optical Correlator (FARCO) bol vyvinutý na Japan Women University na Faculty of Science [3]. Tento systém má tri rôzne konfigurácie, ktoré závisia od rýchlosti rozpoznávania. FARCO je hybridný systém, ktorý integruje technológiu optickej korelácie a digitálnej databázy. Systém môže byť aplikovaný na niekoľko stoviek snímok s operačnou rýchlosťou od 1000 do 5000 spracovaných tvári za sekundu. Pre zvýšenie výkonnosti bol vylepšený na systém S-FARCO, ktorý využíva holografickú pamäť. Takéto vykonanie optickej korelácie bez dekódovania informácie z databázy redukuje čas na spracovanie informácie a teda aj vykonanie korelácie. Na Obr. 5 je zobrazená bloková schéma systému FARCO.



Obr.5 Bloková schéma systému FARCO [3].

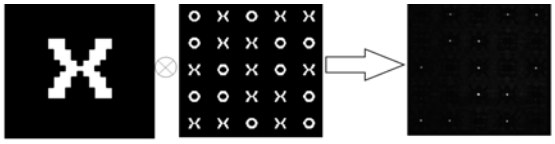
B. 256x256 binárny optický korelátor

256x256 binárny optický korelátor využíva dva SLM na vstup vstupných dát. Bol skonštruovaný Kippom A. Bauchertom a Stevenom A. Seratim v Boulder Nonlinear Systems [4]. Binárne SLM dokážu modulovať svetlo binárne, amplitúdovo alebo fázovo. Na Obr. 6 je zobrazená bloková schéma a fotografia realizovaného 256x256 binárneho korelátora.



Obr.6 Bloková schéma a fotografia 256x256 binárneho korelátora [4].

Teoreticky je načítanie plného rozlíšenia SLM (256x256) možné dosiahnuť za 25 μs. Každé SLM je pripojené na banku 512-tich obrazov. Ako kamera zachytávajúca korelačnú rovinu je použitý CCD snímač Dalsa 256x256. Experiment s týmto korelátorom je na Obr. 7, kde je vyhľadávaný symbol X v poli symbolov X a O.



Obr.6 Experimenty s 256x 256 binárnym korelátorom [4].

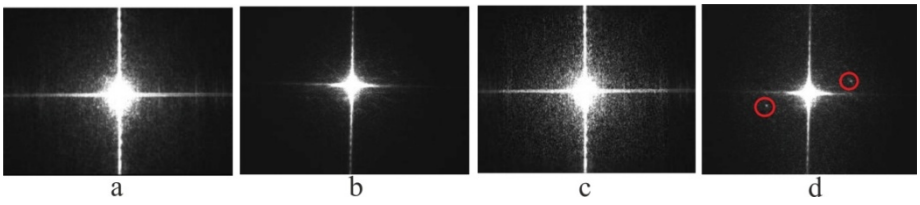
C. Hybridný Optoelektronický korelátor so spojenou transformáciou

Hybrid Optoelectronic Joint Transform Correlator (HOJTC) bo vyvinutý na Chanchung University of Science and Technology v Chanchung v Číne [5]. Tento OC patrí do skupiny korelátorov so spojenou transformáciou. Je navrhnutý na kvalitnú detekciu a rozpoznanie obrazov v scénach s mnohými irelevantnými informáciami vo vstupnom obraze. Korelácia má teda zvýšený odstup signálu od šumu (korelačných špičiek od DC zložky), čo zabezpečuje detekciu objektov aj vo veľmi zašumených vstupných scénach. V HOJTC je požitý Ar^+ laser s výkonom 150mW a vlnovou dĺžkou 514 nm. Ako SLM sú v tomto v korelátore použité dva elektronicky adresovateľné displeje z tekutého kryštálu, EALCD, s rozlíšením 800 x 600 obrazových prvkov. Vlastnosti HOJTC sú predstavené na nasledujúcom experimente. Vstupný aj referenčný obraz sú spojené do jedného. Vstupná scéna je zobrazená na Obr. 7.



Obr.7 Vstupná scéna HOJTC [5].

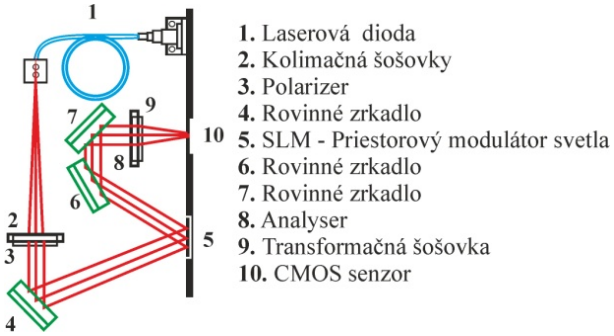
Na Obr. 8 je zobrazené JPS bez spracovania (Obr. 8a) a výstup korelátora získaný z tohto spektra (Obr. 8b), spolu so spracovaným JPS pomocou Waveletovej transformácie (Obr. 8c) a výstup korelátora získaný z takto spracovaného JPS (Obr. 8d). Pri porovnaní na Obr. 8b a Obr. 8d je vidieť že pri rovnakej vstupnej scéne boli získane jasné korelačné špičky, keď bolo JPS spracované pomocou Waveletovej transformácie.



Obr.8 JPS a korelačné roviny HOJTC [5].

D. Cambridge Correlator

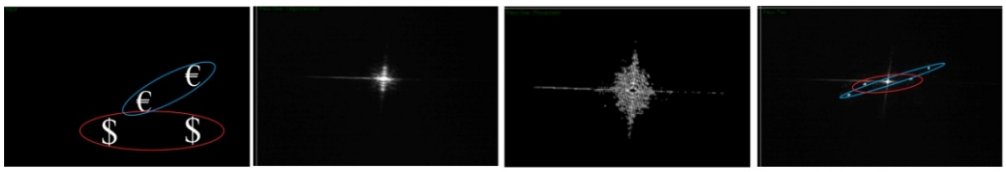
Aj Cambridge Correlator (CC) patrí do skupiny korelátorov so spojenou transformáciou [2]. Bol vytvorený v spolupráci Cambridge University s firmou Cambridge Correlators Limited. Toto zariadenie spolu so softvérom Fourier Optics Experimenter umožňuje študentom aj výskumníkom praktické využívanie optickej Fourierovej transformácie a optickej korelácie na rôznych vstupných scénach. Základné časti optického systému CC sú zobrazené na Obr. 9.



Obr.9 Základné časti CC [2].

Priestorový modulátor svetla SLM je uložený v prednej ohniskovej rovine systému šošoviek a je pripojený k počítaču (PC, laptop) prostredníctvom DVI rozhrania. Obraz na SLM je osvetlený laserovým lúčom z laserovej diódy. Svetlo je modulované obrazmi na SLM, a tak premietané prostredníctvom šošoviek. Tento systém šošoviek realizuje optickú Fourierovu transformáciu v zadnej časti ohniskovej roviny a výstupná intenzita svetla je zachytená CMOS

senzorom. Kolimačné a transformačné šošovky sú umiestnené v presne určených pozíciách, no pre jemnejšie nastavenie je ich možné doladiť. To isté platí aj pre rovinné zrkadlá, ktoré sa pomocou mikro-pohybu nastavovacích skrutiek dokážu dolaďovať pre dosiahnutie čo najlepších výsledkov. Na Obr. 10 sú znázornené vlastnosti korelátorov so spojenou transformáciou a teda aj CC. V tomto experimente boli porovnávané dve dvojice zhodných obrazov. Z tohto experimentu je vidieť, že zhoda je detegovaná iba medzi rovnakými objektmi. Na Obr. 10a je zobrazená vstupná rovina korelačného procesu. Na Obr. 10b je nespracované a na Obr. 10c prahovo spracované Spojené výkonové spektrum (JPS). Obr. 10d zobrazuje výstupnú korelačnú rovinu. V korelačnej rovine vidíme, že v jednom korelačnom procese je možne vykonať viacero korelácií, čo môže rapídne znižovať čas detekcie, resp. rozpoznávania objektov. Korelačné špičky reprezentujú zhodu len medzi rovnakými objektmi.



Obr.10 Porovnanie času dvojrozmernej FT.

CC opísaný v tejto časti je využívaný v LOS KEMT FEI TUKE k vývoju nových systémov na rozpoznávanie dopravných značiek. Tento typ korelátora umožňuje optické spracovanie obrazovej informácie v reálnom čase, čo je pri rozpoznávaní dopravných značiek veľkou výhodou. Ako je vidieť z experimentu na Obr. 10, OC dokáže vykonať viacero optických korelácii súčasne, čo je možné využiť pri detekcii značiek rôznych tvarov, keďže takýto systém nie je viazaný na značky jedného tvaru ako to je doposiaľ u väčšiny rozpoznávacích systémov založených na tvarovej segmentácii.

V. ZÁVER

V tomto príspevku je uvedený prehľad prakticky realizovaných a skúmaných OC. V úvode sú uvedené základné teoretické poznatky k pochopeniu princípu činnosti OC založených na Fourierovej transformácii. V ďalšej kapitole sú popísané dva typy OC (MFC a JTC). Každý typ poskytuje v podstate rovnaké výsledky, líšia sa len spôsobom realizácie optickej korelácie. Každý z nich má svoje výhody aj nevýhody a na základe týchto vlastností je možné určiť najvýhodnejšie pole pôsobností. V ďalšej kapitole boli predstavené niektoré už realizované OC, z každého druhu dva. Spolu s návrhom a realizáciou je uvedené aj pole aplikácie daného korelátora. Optické korelátory majú vo všeobecnosti široké pole pôsobnosti. Dokážu vo veľmi rýchlom čase detekovať objekt v databáze podobných objektov (rozpoznávanie tvári - FARCO), alebo v reálnom čase nájsť zhodu medzi dvoma obrazmi (Cambridge Correlator).

POĎAKOVANIE

Podporujeme výskumné aktivity na Slovensku/ Projekt je spolufinancovaný zo zdrojov EÚ. Tento článok bol vypracovaný v rámci projektu "Centrum excelentnosti integrovaného výskumu a využitia progresívnych materiálov a technológií v oblasti automobilovej elektroniky", ITMS 26220120055 (50%). Táto práca bola taktiež podporovaná a realizovaná z projektu Rozvoj Centra informačných a komunikačných technológií pre znalostné systémy (ITMS kód: 26220120030) na základe podpory operačného programu Výskum a vývoj financovaného z Európskeho fondu regionálneho rozvoja (50%).

LITERATÚRA

[1] J. W. Goodman, *Introduction to Fourier Optic.* Colorado: Roberts & Company ISBN 0-9747077-2-4, 2005.
[2] Cambridge Correlators, :“http://www.cambridgecorrelators.com,“ April 2012.
[3] K. Kodate, E. Watanabe,:“Compact Parallel Optical Correlator for Face Recognition and its Applcation,” Japan 2007.
[4] K. Bauchter, S. A. Serati,: “Data flow architecture for high – speed opticals processors,” *SPIE Proceedings* vol. 3386, 1998.
[5] W. Wang, Y. Chen, C. Liang, H. Miao, “Hybrid optoelectronic joint transform correlator for the recognition of target in cluttered scenes,” *SPIE Proceedings,* vol. 5642, 1998.

Modul senzora teploty a vlhkosti s integrovaným rozhraním 1-wire

Peter HUSNAJ, Miloš DRUTAROVSKÝ

Katedra elektroniky a multimediálnych telekomunikácií, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

peter.husnaj@student.tuke.sk, milos.drutarovsky@tuke.sk

***Abstrakt* — Príspevok opisuje hardvérový a softvérový návrh modulu na meranie teploty a vlhkosti s integrovaným komunikačným rozhraním 1-wire a s možnosťou programovania priamo v aplikácii (In Application Programming) pomocou signálov, protokolu zbernice 1-wire a špecializovaného zavádzacieho programu (bootloadera). Jadro modulu tvorí nízko-príkonový mikroprocesor PIC12LF1822 firmy Microchip, meranie teploty a vlhkosti zabezpečuje senzor SHT21 firmy Sensirion s komunikačným rozhraním I^2C (Inter-Integrated Circuit). 1-wire komunikácia je zabezpečená hardvérovou emuláciou pomocou PIC12LF1822 mikroprocesora.**

***Kľúčové slová* — 1-wire, komunikácia, mikroprocesor, modul, program,**

I. Úvod

V súčasnej dobe stále sa rozvíjajúcich rôznych elektronických súčiastok a zariadení sa vyvíjajú rôzne spôsoby prepojenia týchto zariadení medzi sebou s účelom vzájomnej komunikácie a rovnakého použitého štandardu. Pri vývoji sa kladie dôraz najmä na čo najvyššiu komunikačnú rýchlosť, odolnosť voči rušeniu a zníženie počtu signálov potrebných na zabezpečenie komunikácie. S počtom signálov potrebných na komunikáciu súvisí počet použitých vodičov, ktoré tvoria podstatnú časť finančných nákladov hlavne pri komunikácii na dlhšie vzdialenosti. Komunikačným protokolom, ktorý kladie dôraz na použitie čo najmenšieho počtu vodičov a s tým súvisiacich finančných nákladov je komunikačný protokol 1-wire [1], ktorý využíva na komunikáciu iba jeden signálový vodič.

V príspevku je opísaný návrh modulu teplomera a vlhkomera, ktorý využíva komunikačný protokol 1-wire a je tiež opísané testovanie vytvoreného modulu na dlhej zbernici. Programové vybavenie modulu je navrhnuté tak, aby umožnilo ľahkú modifikáciu aplikačného softvéru priamo v aplikácii len s využitím rozhrania 1-wire.

Návrh podobných zariadení je možné nájsť na internete, avšak žiadne z týchto zariadení nespĺňalo nami požadované vlastnosti, medzi ktoré patria najmä malé rozmery (cca 7x26 mm po finálnej miniaturizácii, ktorá však nie je v tomto príspevku opísaná), nízka spotreba v neaktívnom ako aj v aktívnom režime a možnosť zmeny softvéru priamo v aplikácii (IAP – In Application Programming). Vytvorený modul senzora využíva nízko-príkonový mikroprocesor PIC12LF1822 firmy Microchip [2], v ktorom je emulované kompletné 1-wire komunikačné rozhranie a na meranie teploty a vlhkosti je využitý senzor SHT21 [3].

II. Základné požiadavky navrhovaného modulu senzora

Emulácia 1-wire rozhrania [1] pomocou mikroprocesora je oblasť, ktorá nie je veľmi rozšírená, pretože spoločnosť Maxim vyrába veľké množstvo zariadení s najrôznejšími funkciami, ktoré komunikujú pomocou 1-wire rozhrania, a ktoré uspokoja potreby veľkej väčšiny záujemcov. Napriek tomu existuje potreba vytvoriť zariadenie so špecifickými funkciami ako napríklad zariadenie pomocou, ktorého je možné súčasné meranie teploty a vlhkosti, a ktoré je opísané v tomto príspevku. Výhodou vlastnej realizácie zariadenia komunikujúceho cez 1-wire rozhranie je tiež jednoduchá zmena funkčnosti, resp. pridanie nových funkcií zmenou aplikačného programu mikroprocesora.

Medzi základné vlastnosti, ktorými by mal zhotovený modul na meranie teploty a vlhkosti disponovať patria:

- komunikácia cez 1-wire rozhranie,
- veľmi malý príkon v neaktívnom stave,
- nízky príkon v aktívnom stave (pri meraní),
- činnosť pri nízkom napätí v rozsahu 3 V – 5 V,

- odolnosť voči zvýšenému napätiu napájacieho aj signálového vstupu,
- malé rozmery použitím SMD (Surface-Mount Devices) súčiastok veľmi malých rozmerov,
- možnosť zmeny aplikačného softvéru modulu priamo v zapojení cez 1-wire zbernicu,
- optická indikácia pri meraní a programovaní,
- spätná kompatibilita s teplotným senzorom DS18B20 [4],
- pracovná teplota v rozsahu -40°C až +125°C.

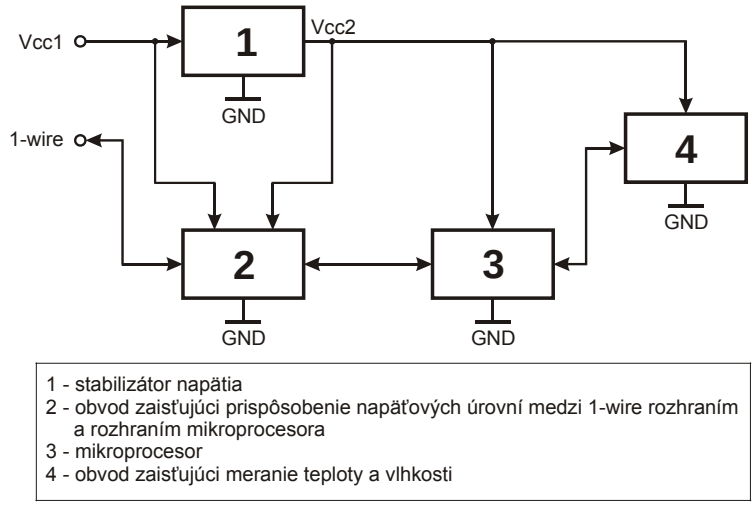
III. HARDVÉROVÝ NÁVRH MODULU SENZORA

Cieľom hardvérového návrhu bolo navrhnúť schému modulu, výber vhodných komponentov pre finálny modul a kompletné overenie ich funkčnosti. Návrh dosky plošných spojov pre miniaturizáciu modulu je cieľom bakalárskej práce [5].

Medzi hlavné komponenty, z ktorých modul senzora pozostáva patria:

- obvod zabezpečujúci napájanie aktívnych prvkov modulu,
- obvod na prispôsobenie napäťových úrovní,
- obvod na emuláciu 1-wire rozhrania,
- obvod senzora na meranie teploty a vlhkosti.

Bloková schéma navrhnutého modulu senzora je zobrazená na Obr. 1.



Obr. 1 Bloková schéma modulu senzora teploty a vlhkosti s emulovaným rozhraním 1-wire

Pri výbere vhodných obvodov pre modul senzora sa kládol dôraz na malé rozmery obvodu, nízku spotrebu, schopnosť všetkých obvodov pracovať pri teplote v rozmedzí -40°C až +125°C.

Ako obvod na meranie teploty a vlhkosti bol zvolený moderný obvod SHT21 [3] spoločnosti Sensirion. Obvod SHT21 umožňuje merať teplotu v rozsahu -40 °C až +125 °C a vlhkosť v rozsahu 0 % až 100 %. Obvod je možné objednať vo viacerých verziách v závislosti od komunikačného protokolu, ktorým môže byť: I^2C (Inter-Integrated Circuit) protokol, PWM (Pulse-Width Modulation) alebo analógový Sigma-Delta modulovaný formát. Kvôli jednoduchej komunikácii s mikroprocesorom bol pre navrhovaný modul senzora zvolený I^2C protokol. Obvod je vyrábaný v puzdre DFN (Dual-Flat No-leads), ktorého rozmery sú iba 3x3 mm. Spotreba obvodu je veľmi malá, pri aktívnom režime 330 µA (VDD = 3 V, T < 60 °C) a v neaktívnom režime iba 0,4 µA (VDD = 3 V, T < 60 °C).

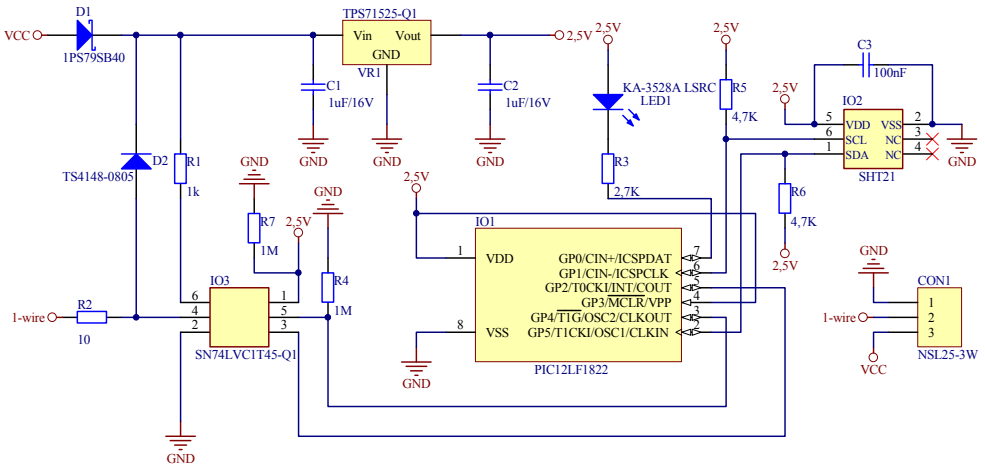
Najdôležitejšou časťou celého modulu je mikroprocesor, v ktorom je emulované komunikačné rozhranie 1-wire a taktiež riadená komunikácia s obvodom SHT21. Keďže 1-wire komunikácia používa signály s dĺžkou radovo µs, dôraz pri výbere mikroprocesora bol kladený na čo najvyššiu taktovaciu frekvenciu mikroprocesora s použitím vnútorného oscilátora, aby mal výsledný modul senzora čo najmenšie rozmery. Pre modul bol zvolený nízko-príkonový mikroprocesor PIC12LF1822 [2] od spoločnosti Microchip, ktorého taktovaciu frekvenciu vnútorného oscilátora je možné nastaviť až na hodnotu 32 MHz. Dôležitou vlastnosťou tohto mikroprocesora je podpora reprogramovania v aplikácii (IAP) a ochrana určitej oblasti programovej pamäte pred týmto reprogramovaním, čo bolo využité pri zmene aplikačného programu priamo cez 1-wire rozhranie s využitím vyvinutého zavádzacieho softvéroveho kódu. Pre cieľový modul bol vybraný mikroprocesor PIC12LF1822 v puzdre

DFN podobne ako senzor SHT21, pričom bola využitá verzia s pracovnou teplotou v rozmedzí -40 °C až +125 °C.

Keďže 1-wire zbernica je napájaná napätím +5 V a obvod SHT21 je možné napájať maximálnym napätím 3,6 V [3] bolo potrebné prispôsobiť napájacie napätie celého modulu. Pri prispôsobovaní sa kládol najmä dôraz na minimalizáciu spotreby celého modulu, preto bolo napájacie napätie zvolené čo najmenšie, pri ktorom boli schopné pracovať všetky obvody modulu. Pri voľbe napájacieho napätia sa taktiež zohľadňovala maximálna taktovacia frekvencia mikroprocesora, ktorá klesá so znižujúcim sa napájacím napätím. Pre správnu funkčnosť obvodu SHT21 je potrebné napájacie napätie minimálne 2,1 V. Pri tomto napätí je schopný pracovať aj mikroprocesor, preto bolo zvolené najbližšie vyššie štandardne používane napájacie napätie a to napätie s hodnotou 2,5 V, pri ktorom mikroprocesor pracuje v celom rozsahu taktovacích frekvencií a teplôt. Na konverziu vstupného napätia na napätie s hodnotou 2,5 V bol zvolený obvod stabilizátora TPS71525-Q1 [6] od spoločnosti Texas Instruments, ktorého hlavnými prednosťami sú nízka spotreba (3,2 µA pri odoberanom výstupnom prúde 50 mA), malé puzdro SOT-323 (DCK) s rozmermi 2,15x2,3 mm a možnosť použiť keramické kondenzátory s nízkou kapacitou (≥ 0,47 µF) na výstupe pre stabilizáciu vnútorného riadenia obvodu. Obvod je schopný pracovať v teplotnom rozsahu -40 °C až +125 °C a pri napätí min. 2,5 V a max. 24 V.

Modul teplomera a vlhkomera je napájaný napätím s veľkosťou 2,5 V, preto nie je možné pripojiť mikroprocesor, ktorý emuluje komunikačné rozhranie 1-wire, priamo k 1-wire zbernici pracujúcej s napätím max.5 V. Priamym pripojením by mohlo dôjsť k poškodeniu periférii mikroprocesora. Keďže 1-wire komunikácia je založená na obojsmernej komunikácii po jednom vodiči bolo potrebné prispôsobiť napäťové úrovne medzi 1-wire vodičom zbernice a vývodom mikroprocesora. Pri prispôsobení sa zohľadňovala aj vstupná kapacita modulu a spotreba modulu z 1-wire dátového vodiča. Na prispôsobenie napäťových úrovní bol zvolený špecializovaný obvod – napäťový translátor SN74LVC1T45-Q1 [7] od spoločnosti Texas Instruments, ktorý bolo možné objednať ako testovaciu vzorku avšak iba do max. pracovnej teploty 105 °C. V prípade použitia vo vyšších teplotách ako 105 °C je potrebné modul osadiť obvodom SN74LVC1T45-EP pracujúcim až do 125 °C. Obvod obsahuje vstup DIR [7], pomocou ktorého je ovládaný smer vstupu/výstupu. Obvod je vyrábaný v rovnako malom puzdre ako stabilizátor TPS71525-Q1 [6].

Celková schéma zapojenia modulu teplomera a vlhkomera s použitím uvedených obvodov je zobrazená na Obr. 2.



Obr. 2 Celková schéma zapojenia modulu teplomera a vlhkomera

Rezistor R2 chráni vstupnú časť obvodu pred zvýšením vstupným napätím. Hodnota rezistora R2 bola zvolená ako kompromis medzi čo najlepšou ochranou signálového vstupu a zároveň čo najmenším skreslením vstupného signálu a taktiež malou hodnotou odporu v zopnutom stave. Úlohou diódy D2 je korektné napájanie elektroniky, hlavne mikroprocesora a obvodu SN74LVC1T45-Q1 aby sa tieto obvody nachádzali v definovanom stave pri náhodnom výpadku napájacieho napätia, prípadne pri nábehu napätia. Tým sa zabezpečí zníženie špičkovej spotreby celého modulu pri nábehu zbernice. Rezistor R1 chráni napájaciu časť obvodu SN74LVC1T45-Q1 pred zvýšeným napájacím napätím a nebráni v správnej funkcii obvodu, keďže obvod má veľmi malú spotrebu v režime kedy je vývod DIR v log. 0, čím na rezistore R1 vzniká veľmi malý úbytok napätia. V režime kedy je vývod DIR v log. 1 a vstup A je taktiež v log. 1 sa obvod nepoužíva, log. 1 na výstupe B zabezpečuje zdvíhací

rezistor 1-wire zbernice pri uvedení vstupu A do stavu vysokej impedancie. LED1 slúži na indikáciu pri meraní teploty a taktiež pri vstupe modulu do programovacieho režimu. LED1 bolo potrebné vybrať s čo najmenším úbytkom napätia a prúd, ktorý ňou preteká nastaviť na najmenšiu hodnotu, pri ktorej je jas LED (Light Emitting Diode) dostatočný. Tým klesne celková spotreba celého modulu. Pri zostavení celého testovacieho modulu na univerzálnej doske s plošnými spojmi bola použitá vývodová LED dióda s nízkym príkonom L-934 LID, ktorej prúd bol experimentálne nastavený na hodnotu 320 µA pomocou rezistora R3 s veľkosťou 2,7 kΩ. Opis funkcie ostatných súčiastok použitých v schéme zapojenia je možné nájsť v diplomovej práci [8].

LED dióda je jediná súčiastka v module senzora, ktorá pracuje iba v teplotnom rozsahu -40 °C až +85 °C a teda nie je možné ju použiť v prostrediach, kde meraná teplota presahuje +85 °C. V prípade, že bude modul používaný v takýchto prostrediach nesmie byť LED dióda v module osadená čím sa stratí možnosť indikácie avšak funkčnosť modulu ostane zachovaná až do teploty +125 °C.

IV. SOFTVÉROVÝ NÁVRH MODULU SENZORA

Cieľom softvérového návrhu bolo vytvoriť zdrojový kód v jazyku C pre použitý mikroprocesor PIC12LF1822 [2]. Pri vytváraní zdrojového kódu bolo použité vývojové prostredie MPLAB v.8.84 od spoločnosti Microchip a pri kompilácií zdrojového kódu bola využitá testovacia verzia kompilátora CCS C v4.122, ktorú je možné integrovať do prostredia MPLABu.

Na začiatku vývoja kódu bola využívaná maximálna taktovacia frekvencia mikroprocesora 32 MHz. Na zníženie spotreby celého modulu bolo potrebné využiť nízko-príkonový (sleep) režim mikroprocesora, v ktorom je okrem iných obvodov odpájané aj časti oscilátora generujúce taktovacie frekvencii väčšie ako 16 MHz a PLL (Phase Lock Loop) obvody. Taktovacia frekvencia 32 MHz však využíva PLL obvody, ktorých nábeh pri prechode z nízko-príkonového do aktívneho režimu trvá podľa katalógového listu mikroprocesora max. 2 ms ([2], str. 357). Tento čas je príliš dlhý, keďže 1-wire komunikácia využíva časové sloty rádovo v mikrosekundách a teda nie je možné použiť túto taktovaciu frekvenciu mikroprocesora pri súčasnom použití nízko-príkonového režimu. Kompromisom medzi použitím čo najvyššej taktovacej frekvencie a súčasným použitím nízko-príkonového režimu sa stala frekvencia 16 MHz. Táto frekvencia je v poradí druhou najvyššou frekvenciou, ktorú je možné zvoliť, pričom nevyužíva PLL obvody. Využíva vnútorný oscilátor s vysokou frekvenciou, ktorého nábeh pri prechode z nízko-príkonového do aktívneho režimu trvá maximálne 8 µs (typicky 5 µs) ([2], str. 357) čo je postačujúci čas.

Veľkou výhodou navrhnutého modulu senzora je využitie princípu zavádzacieho programu [9]. Nami vyvinutý špecializovaný zavádzač umožňuje zmenu alebo aktualizovanie zdrojového kódu aplikačného programu priamo pomocou 1-wire rozhrania. Celý zdrojový kód pre mikroprocesor je teda rozdelený na zavádzací program a aplikačný program. Zavádzací program je uložený v programovej pamäti od adresy 0 po adresu 511 čo je najmenšia časť pamäte FLASH, ktorú je možné hardvérovo chrániť (nastavením konfiguračných bitov pri programovaní mikroprocesora) pred vlastným IAP reprogramovaním, preto bola pre zavádzací program zvolená práve táto oblasť.

Príjem a vysielanie dát cez 1-wire rozhranie je riešené v podprograme prerušenia. Prerušenie je pomocou špeciálnych funkčných registrov nastavené tak, aby bolo vyvolané dobežnou hranou na vývode RA2. Prerušenie používa tzv. vlastného dispečera, čo je kód generovaný kompilátorom, ktorý je generovaný pred vstupom do prerušenia. Výhoda použitia vlastného dispečera spočíva v zmenšení dĺžky kódu pre obsluhu prerušení generovaného kompilátorom a teda aj vo zvýšení rýchlostí reakcie na vyvolané prerušenie. Takýmto spôsobom sa podarilo skrátiť skok z hlavného programu do prerušenia z pôvodných 4,75 µs na 0,75 µs, čo je cca 6-krát rýchlejšie ako pri použití kompilátorom generovaného všeobecného dispečera. Pri použití pôvodného času bolo veľmi problematické aplikačným kódom dodržať podmienky časovania 1-wire zbernice, ktorá je z pohľadu časovania mikroprocesora PIC12F1822 s taktom 16 MHz veľmi rýchla. 1-wire komunikácia riešená v prerušení sa nachádza v časti pamäte pre zavádzací program. Niektoré funkcie použité v zavádzacom programe sú z dôvodu efektívneho využitia programovej pamäte aplikačného programu, ktorej veľkosť je 1512 slov, zdieľané medzi zavádzacím a aplikačným programom, čo znamená, že funkcie sa nachádzajú v zavádzacom programe, ale sú volané (aj) z aplikačného programu.

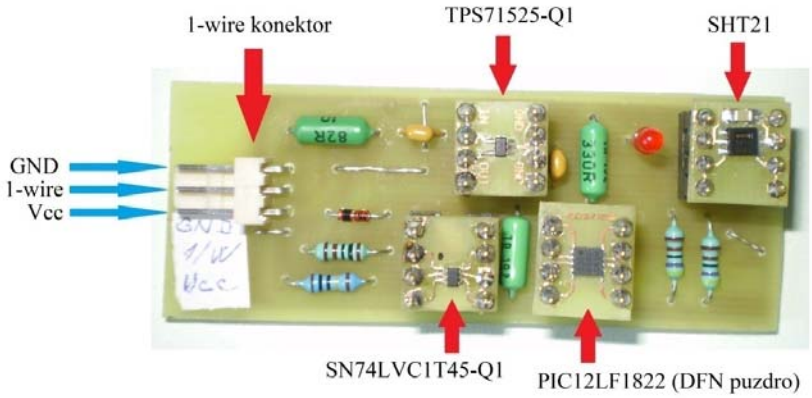
Na začiatku vývoja mal zdrojový kód zavádzacieho programu veľkosť výrazne presahujúcu 512 slov a bolo veľkou výzvou upraviť tento kód na veľkosť menšiu ako 513 slov, ktorá je potrebná na to, aby bolo možné kód vložiť do programovej pamäte od adresy 0 až po adresu 511 a túto oblasť potom chrániť pred vlastným IAP reprogramovaním. Preto boli vykonané rôzne optimalizácie obsluhy prerušenia a hlavného programu medzi, ktoré patrilo najmä

použitie niektorých vysoko optimalizovaných funkcií v jazyku Assembler pre výpočty kontrolných súčtov CRC8 a CRC16 a použitie 8-bitových smerníkov v zavádzacom programe namiesto pôvodných 16-bitových, ktoré sú využívané už len v aplikačnom programe.

Aplikačný program je umiestnený v programovej pamäti mikroprocesora v oblasti adries 512 – 2047 a jeho úlohou je meranie teploty a vlhkosti z pripojeného senzora SHT21 [3] a prenos týchto údajov vo vhodnom formáte cez 1-wire rozhranie. Aplikačný program je navrhnutý tak, aby sa celý modul senzora v čo najväčšej miere z hľadiska komunikácie podobal senzoru DS18B20 [4]. V aplikačnom programe sú implementované všetky príkazy s výnimkou príkazu *READ POWER SUPPLY*. Podporované 1-wire príkazy modulu sú: *SEARCH ROM*, *READ ROM*, *MATCH ROM*, *SKIP ROM*, *ALARM SEARCH*, *CONVERT TEMPERATURE*, *WRITE SCRATCHPAD*, *READ SCRATCHPAD*, *COPY SCRATCHPAD*, *RECALL EEPROM* a *WRITE HUMIDITY SCRATCHPAD*. Všetky tieto príkazy s výnimkou posledného príkazu *WRITE HUMIDITY SCRATCHPAD* sú podporované taktiež obvodom DS18B20, čím je zabezpečená spätná kompatibilita s týmto obvodom. Ich funkciu a opis je možné nájsť v katalógovom liste obvodu DS18B20 [4]. Posledný príkaz je doplnený pre prácu s užívateľským registrom senzora SHT21 a prahmi pre vlhkosť.

V. Praktická realizácia a testovanie modulu senzora

Pre navrhnutú schému zapojenia modulu senzora bola kvôli testovaniu všetkých cieľových súčiastok pred výslednou miniaturizáciou modulu a tiež testovaniu zavádzacieho a aplikačného programu navrhnutá testovacia doska plošných spojov (Obr. 3), ktorá je súčasťou bakalárskej práce [5]. Miniaturizácia modulu nie je cieľom príspevku a bude opísaná v ďalších príspevkoch po dokončení finálnej verzie modulu.



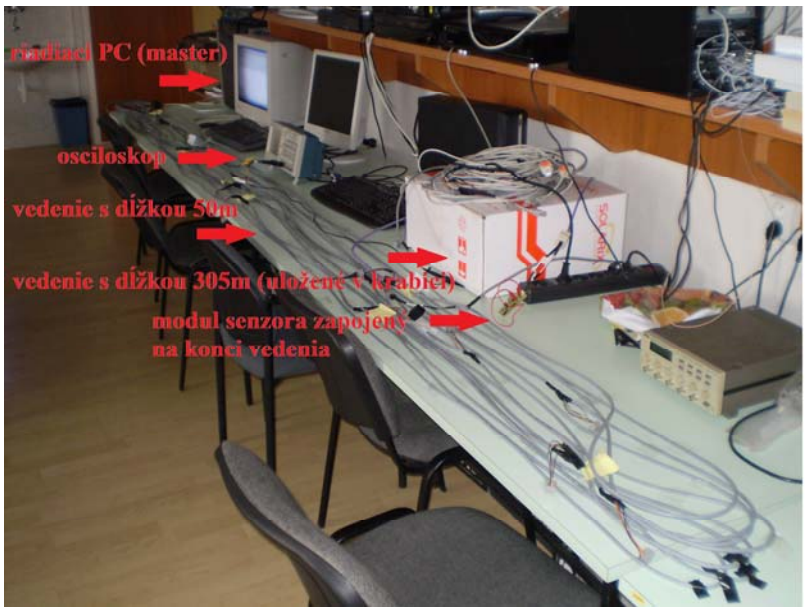
Obr. 3 Testovacia verzia modulu senzora

Pred testovaním modulu bol do mikroprocesora PIC12LF1822 s využitím štandardného programátora nahratý kód zavádzacieho programu. Podmienkou vstupu zavádzacieho programu do programovacieho režimu je vygenerovanie špeciálneho signálu na 1-wire zbernici, čím je zabezpečené, že pri bežnej prevádzke je veľmi malá pravdepodobnosť vstupu modulu senzora do tohto režimu. Bežné 1-wire adaptéry, napr. USB–1-wire adaptér [10] však nedokážu vygenerovať tento špeciálny signál, preto bol na programovanie modulu senzora cez 1-wire rozhranie vytvorený špecializovaný programátor, ktorého úlohou je čítanie Intel-HEX súboru z počítača cez sériové rozhranie a vysielanie prijatých dát cez 1-wire rozhranie k pripojenému modulu senzora. Na testovanie v spojení s 1-wire adaptérom bol využitý program 1-wire Viewer, ktorý spolupracuje so všetkými 1-wire zariadeniami spoločnosti Maxim a teda je ho možné použiť aj s modulom senzora, keďže tento je spätne kompatibilný so senzorom DS18B20. Na testovanie bol taktiež použitý program vytvorený na Katedre elektroniky a multimediálnych telekomunikácií, ktorým bolo možné testovať aj meranie vlhkosti a zápis prahov pre vlhkosť.

Modul bol počas vývoja testovaný priamym pripojením k USB–1-wire adaptéru [10] pomocou kábla s dĺžkou niekoľko centimetrov a taktiež na vedení s dĺžkou cca 355 m (Obr. 4).

Pri meraní celkovej spotreby modulu senzora boli pri teplote 22 °C namerané tieto hodnoty:

- prúd v neaktívnom režime → 11 µA,
- prúd v aktívnom režime bez prebiehajúcej 1-wire komunikácie → 917 µA,
- špičkový prúd pri meraní teploty a rozsvietenej LED dióde → 2,098 mA.



Obr. 4 Pracovisko pri testovaní modulu na dlhom vedení

VI. ZÁVER

Navrhnuté hardvérové a softvérové riešenie modulu senzora je výstupom diplomovej práce [8]. Modul senzora je navrhnutý pre použitie v priemyselnom prostredí s pracovnou teplotou v rozsahu -40°C až +125°C. Modul senzora bol testovaný na 1-wire zbernici pomocou nato vytvoreného softvérového vybavenia. Modul bol otestovaný aj v kritických podmienkach dlhej zbernice (cca 355 m), pričom pracoval spoľahlivo. Na programovanie modulu senzora priamo v aplikácii cez 1-wire rozhranie bol navrhnutý a zrealizovaný špecializovaný programátor. Moduly boli zrealizované a otestované na testovacej doske plošných spojov s cieľovými obvodmi a po zrealizovaní finálnej miniaturizovanej verzie budú pripravené na použitie.

POĎAKOVANIE

Táto publikácia vznikla vďaka podpore v rámci operačného programu Výskum a vývoj, pre projekt: Kompetenčné centrum znalostných technológií pre inovácie produkčných systémov v priemysle a službách, kód ITMS: 26220220155, spolufinancovaný zo zdrojov Európskeho fondu regionálneho rozvoja.

LITERATÚRA

[1] Springbok Digitronics: Understanding 1-wire Series [online]. Springbok Digitronics, 2004. Aktualizované 19.08.2004. Dostupné na internete: <http://www.1wire.org/Files/Articles/1-Wire-Design%20Guide%20v1.0.pdf>.
[2] Microchip: PIC12F/LF1822 Data Sheet [online]. Microchip, 2010. Aktualizované 08.04.2010. Dostupné na internete: <http://ww1.microchip.com/downloads/en/DeviceDoc/41413B.pdf>.
[3] Sensirion: Datasheet SHT21 [online]. Sensirion, 2011. Aktualizované December 2011. Dostupné na internete: <http://www.sensirion.com/en/pdf/product_information/Datasheet-humidity-sensor-SHT21.pdf>.
[4] Maxim: Programmable Resolution 1-Wire Digital Thermometer [online]. Maxim, 2008. Aktualizované 2008. Dostupné na internete: <http://datasheets.maxim-ic.com/en/ds/DS18B20.pdf>.
[5] S. Slovák, "Senzor vlhkosti s rozhraním 1-wire". Bakalárska práca, Košice, TU FEI, 2012.
[6] Texas Instruments: TPS71525-Q1 [online]. Texas Instruments, 2004. Aktualizované August 2011. Dostupné na internete: <http://www.ti.com/lit/ds/symlink/tps71525-q1.pdf>.
[7] Texas Instruments: Single-bit Dual-supply Bus Transceiver with Configurable Voltage Translation and 3-State Outputs [online]. Texas Instruments, 2006. Aktualizované August 2009. Dostupné na internete: <http://www.ti.com/lit/ds/symlink/sn74lvc1t45-q1.pdf>.
[8] P. Husnaj, "Riadiace moduly s rozhraním 1-Wire". Diplomová práca, Košice, TU FEI, 2012.
[9] Your Electronics Open Source: What a Microcontroller Bootloader Is and How It Works? [online]. Dostupné na internete: <http://dev.emcelettronica.com/what-microcontroller-bootloader-and-how-it-works>.
[10] P. Stopko, "Moduly pre senzorovú sieť na báze 1-Wire protokolu". Diplomová práca, Košice, TU FEI, 2007.

MHP aplikácie v interaktívnej televízii

Rastislav KOKOŠKA, Stanislav MARCHEVSKÝ

Katedra elektroniky a multimediálnych telekomunikácií, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

rastko@pobox.sk, stanislav.marchevsky@tuke.sk

***Abstrakt*—V tomto článku sme sa zamerali na služby v oblasti IPTV a konkretnejšie MHP aplikácie. Prispevok poskýtuje prehľad možných typov služieb, ich výhody a jej vlastnosti. MHP služby v interaktívnej TV popisujú architektúru, vrstvy, profily ,využitie a aplikácie MHP. Možné typy MHP služieb umožňujú rôznu interaktivitu využitia IPTV napríklad v LCD TV prijímačoch, ktoré sú spomenuté na konci článku. Spomínané MHP služby majú viacero výhod ako sú napríklad výhody interaktívnosti služieb pre sledovanie TV. Táto interaktivita sa taktiež používa v špecializovaných prijímačoch nazývaných SetBox. Medzi ďalšie žiadané služby koncových užívateľov patria aj služby ako EPG, rýchly pristup na internet a rôzne multimediálne služby.**

V budúcnosti ako aj v prítomnosti sa tieto služby čoraz viac využívajú a sú čoraz viac žiadané či už zo strany poskytovateľov ale aj koncových užívateľov.

***Kľúčové slová* — IPTV, MHP, interaktívna televízia, VAS - Value Added Services, EPG, Application Infomation Table (AIT), JAVA VM (JAVA Virtual Machine), API (application programming interface), DVB-J, Digital Storage Media - Command and Control (DSM-CC), DVB-S- satelitné vysielanie, DVB-T- terestriálne vysielanie, DVB-C- káblové vysielanie, Pay Per View-platené služby, Video On Demand- video na želanie, GEM -Globally Executable MHP.**

I. ÚVOD

V tomto článku predstavíme teoretický základ pre použitie MHP služieb v IPTV a jej poskytované ďalšie služby. Dôvodom nášho záujmu je mobilita služieb, nielen klasické pozeranie TV programov ale jej interaktivita TV programov, sledovanosť a kreatívnosť výberu možných služieb. Na konci článku uvádzame možnosti praktickej implementácie do existujúcich a nových zariadení ako sú TV prijímače a set top boxy (satelitné a káblové digitálne prijímače).

A. iTV a IPTV služby

ITV je televízia s interaktívnym obsahom. Kombinuje sledovanie tradičnej televízie s komunikáciou prostredníctvom spätného kanála, väčšinou prostredníctvom počítačovej siete (napr. internetu). IPTV (Internet Protocol TV), ktorá na distribúciu televíznych a videoslužieb („video na požiadanie“) používa už existujúcu dátovú sieť, väčšinou privátnu a nie verejnú. Predplatitelia tak v "jednom balíku" a pri jedinom pripojení majú k dispozícii telefón, televíziu a on-line služby od jedného poskytovateľa. Z uvedeného vyplýva, že IPTV môže v sebe zahŕňať aj to, čo označujeme ako iTV. Video na požiadanie (VOD - video on demand) môže byť podmnožinou iTV [1]. Interaktívne služby sú jedným z najväčších lákadiel, ktoré digitálna televízia prináša. Samotný výraz „interaktivita“ znamená pre diváka možnosť zmeniť obsah na obrazovke. Typickým predstaviteľom je teletext. [12]

B. Služby s úplnou interaktivitou

Dosiahnuť plnú interaktivitu služieb, znamená potrebu obojsmernej komunikacie medzi uživateľom a poskytovateľom danej služby. Pri dnešných set-top boxoch je pre tvorbu spätného kanálu využívaná predovšetkým technológia ethernet a integrovaný telefónny modem. Ethernet umožňuje začlenenie set-top-boxov do lokálnej siete a následné využitie ktoréhokoľvek ďalšieho pripojenia k internetu (ADSL, ISDN, 802.11). Užívateľ, ktorý ma set-top-box so spätným kanálom, bude môcť napríklad prijímať a odosielať e-mailové správy, prehliadať internetové stránky, prevadzkovať internetové nákupy, sťahovať filmy, hrať počítačové hry atď. [12]. Možnosti obrazových informácií v interaktívnej digitálnej televizii sa výrazne rozširujú a zvyšujú jej funkčnosť. Jej obsah možno rozdeliť do dvoch hlavných skupín. Jednu tvoria TV programy a formáty a druhú služby s pridanou hodnou (**VAS** - Value Added Services)[2].

Služby s pridanou hodnotou môžu súvisieť s programom, ale môžu byť aj samostatné. Najbežnejším typom VAS sú informačné (vrátane elektronického sprievodcu programom-EPG), komunikačné, zábavné, transakčné služby a všeobecne interaktívne TV programy. Z pripravovaných dátových služieb: online hlasovanie počas seriálov, online nakupovanie, home banking, výber kamery pri športových prenosoch, interaktívne hry, prehľadanie webových stránok, plne grafický teletext – superteletext, Pay Per View, Video On Demand, apod.)

Pay Per View - pri platených seriáloch je tá predstava, že v rámci súhrného toku DVB-T bude určitý seriál vysielaný opakovane (v slučke), a divák si prostrednictvom spätného kanálu (využívaný obojsmerne) vyžiada údaje potrebné k jeho dekódovaniu a zaplatí za túto službu. Potom môže príslušný seriál buď sledovať priamo (ako stream), alebo si to nahrať na lokálny pevný disk a pozrieť neskôr, prípadne niekoľko krát – podľa toho, aké práva si kúpil.

Video on Demand – Služba „Video On Demand" (video na želanie) umožňí divákovi zostaviť si vlastný program. Pokiaľ volí v rámci vysielaného dátového toku, môžu to byť krátke videosekvencie, napr. opakované zábery pri sledovaní športových stretnutí. Objemnejší a špecifikovanejší obsah (celé televízne seriály alebo celovečerné filmy) potom bude prenášaný pomocou spätného kanálu len k divákovi, ktorý si ho zvolil [7].

Dnes existuje množstvo rôznych súťaží a show, ktoré na interakciu s divákom používajú SMS ako spätný kanál. Obsah SMS je využitý pri hlasovaní divákov (ovplyvňuje priebeh súťaže alebo zábavného programu), alebo je zobrazený na obrazovke, ako príspevok chatu. Divák sa môže svojimi odpoveďami doma, pri svojom TV prijímači zúčastniť súťaže, môže v priebehu športovej súťaže a pretekov vsadiť na výsledok (on-line stávkovanie).

Interaktívna TV reklama (i-advertising) ponúka rôzne spôsoby interakcie medzi divákom a internetom. Môže ísť o vyžiadanie dodatočnej informácie alebo skúšobnej vzorky pomocou spätného kanála, účasť v rôznych súťažiach usporiadaných interpretmi, prípadne iba o sprostredkovanie spätnej väzby na nich, ale aj o on-line obchodovanie (T-commerce).
Súčasťou možností iTV je aj elektronická komunikácia medzi divákmi, využitie elektronickej pošty a surfovanie v internetovej sieti.Všetko spomínané tvorí nielen technologický, ale aj hodnotový reťazec, ktorého celkové výnosy sú samozrejme veľké a veľkým dielom závisia a budú zavisieť práve na vôli a záujme zákazníkov. [3].

Interaktívna televízia rozširuje prijímanie a spracovanie informácií na viaczdrojové a viacprístrojové, čo znamená použitie nielen televízneho prijímača. Keďže použitie iTV je niečo medzi sledovaním televízie a prácou s počítačom, pričom tieto dve odlišné činnosti sa k sebe približujú a spájajú, vyžaduje si to postupný vývoj nových hardvérových a softvérových platforiem.

II. MHP – MULTIMEDIA HOME PLATFORM

Je to štandard, na ktorom su postavené takmer všetky doplnkové služby digitálneho televízneho vysielania, predovšetkým najpružnejšej (interaktivný) multimediálnej aplikacie (jazyk Java). Bol navrhnutý projektom DVB. Platforma prináša technické riešenia užívateľských terminálov (prijímač DVB-MHP), ktorá umožňuje príjem interaktívnych služieb. Systém umožňuje vylepšené zobrazenia, výber kamery, ktorá sníma zápas, interaktivný prístup (komunikacia z obidvoch stran, ako užívateľa, tak aj druhej strany - serveru), ďalej internetový prístup. [4]

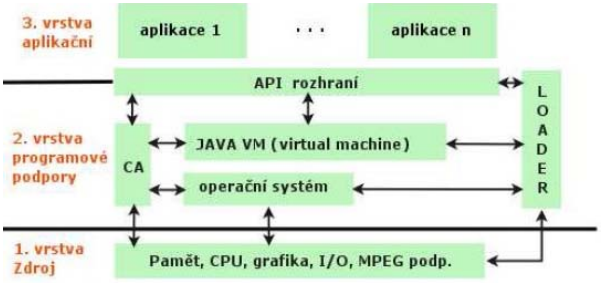
A. Základné vrstvy MHP

Sú to zdroj, programové vybavenie a aplikácie. Hardware nie je definovaný. Zdrojmi sú spracovanie formátu MPEG, procesor, pamäť a grafický systém. Aplikácie navrhnuté pre MHP musia spĺňať stanovené aplikačné rozhranie API (Application Programming Interface), ktoré je definované týmto systémom [5].

Zdroje (hardwarové a softwarové prostriedky) sú potrebné pre spracovanie obrazových a zvukových dát, napr. MPEG-2 dekodér, alebo hlavný a grafický procesor, ktoré umožňujú prístup k dostupným aplikačným rozhraniam.

Systémový software – operačný software, ktorý obsahuje správcu aplikácií a vytvára podporu pre základné transportné protokoly. Súčasťou je Java Virtual Machine, ktorá oddeľuje výrobcom navrhnutý hardware a software od normou definovaného aplikačného rozhrania (API).

Aplikácie, ktoré majú prístup k platforme len cez definované aplikačné rozhranie.[5].



Obr. 1 Základné vrstvy a podrobnejši opis štruktúry MHP [6],[8]

Dôležitou súčasťou je aplikačný manažér, pomocou ktorého sa jednotlivé aplikácie spustia, riadia a ukončujú. Pre aktualizáciu softwaru obsahuje vrstva tzv. „Loader“, ktorá je implementovana podľa výrobcu.

Dôležitou sučástou tejto vrstvy je **JAVA VM (JAVA Virtual Machine),** čo je de facto zapúzdrenie hardwaru pre využitie aplikačnej vrstvy. To znamená, že zadané aplikácie pristupujú k hardwarovým zdrojom, ktoré sú vykonávané cez **JAVA VM**.

Do aplikačnej vrstvy spadajú jednotlivé programy, riadené aplikačným manažérom druhej vrstvy. Tieto aplikácie sa spúšťajú nad vrstvou aplikačného rozhrania **API (application programming interface)**, ktorá má pevne definované pravidlá prístupu a preto umožňuje oddeliť samotnú aplikáciu od nižších vrstiev a dosahuje tak multiplatformnosť[9].

B. Profily - využitia MHP služieb

MHP je možné rozdeliť na tri základne profily.

Rozšírené televízne vysielanie sa týka set-top boxov bez spätného kanála s pasívnou interaktivitou. Plne definovaný je v špecifikácii MHP 1.0 - obsahuje Java VM, DVB-J API, transportné protokoly IP.

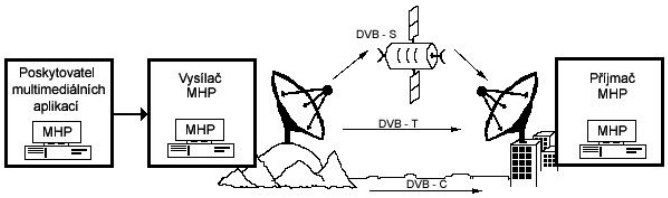
Interaktívna televízia - tiež definovaná v MHP 1.0, set-top boxy ale zahŕňajú spätný kanál, čím umožňujú vyšší stupeň interaktivity, obsahuje rozšírené aplikácie DVB-J API a interaktívne transportné protokoly IP.

Prístup k internetu - set-top boxy s vysokým výpočtovým výkonom, veľkou pamäťou a s najvyšším stupňom interaktivity. Tento profil je definovaný v MHP 1.1, obsahuje Java API pre prístup k internetu, transportné protokoly pre vysielanie, DVB-HTML a ďalšie.

V súčasnosti vychádza aj špecifikácia MHP 1.2, ktorá pridáva podporu pre poskytovanie služieb DVB prostredníctvom širokopásmových sietí (IPTV) s novým API. Tiež pridáva podporu pre prevádzkovateľov dodávaných aplikácií, ktoré bežia po celú dobu behu prostredia MHP.[5].

C. Aplikácie MHP služieb

Na Obr.2 je vidieť blokovú schému prenosového reťazca signálu digitálnej televízie a MHP. Poskytovateľom multimediálnych aplikacií sa rozumie dodávateľ, ktorý tieto aplikácie (napr. počasie, televízny program, krížovky, internet, hry, domáce bankovníctvo, a pod.) vytvára a dáva ju k dispoziícii užívateľovi (divákovi). Vysielanie signálu digitálnej televízie môže prebiehať troma spôsobmi. Buď je to satelitné vysielanie (DVB-S), pozemské (terestriálne) vysielanie (DVB-T) a nakoniec to je káblové vysielanie (DVB-C). Na prijímacej strane je prijímač (set-top box), pomocou ktorého si užívateľ (divák) zvolí tu službu, na ktorú má práve náladu. Set-top box je preto najdoležitějším článkom prenosového reťazca z hľadiska užívateľa, pretože práve na ňom sa spustia multimediálne aplikácie. Prenos multimediálnych aplikácií v štandarde MHP je založený na programovacom rozhraní API (Application Programming Interface). Pre multimedialnú platformu bol zvolený programovací jazyk Java, ktorý je multiplatformny a nezávislý na použitom HW a SW.



Obr. 2 Vysielanie a príjem MHP.

Na túto platformu sú kladené nasledujúce komerčné požiadavky: rozšíriteľnosť, možnosť aktualizácie, interoperabilita, aplikácia musí byť spustiteľná pre rôznych výrobcov hardwaru a poskytovateľov pripojenia, oddelenie dát od aplikácií, otvorenosť štandardu, podpora systému s podmieneným prístupom (platené služby) [5], otvorený a voľne modifikovateľný štandard, vrstvový model MHP platformy, minimálné finančné zaťaženie pre spustenie interaktivných služieb, nezávislosť API a CA (Conditional Access System – jednotka podmieneho prístupu), musí byť umožnená podpora odlišných CA prístupu a Common Interface[10].

Aplikácie určené pre MHP sú preto napísané v jazyku JAVA alebo HTML. Tým sa stávajú nezávislé od hardwarovej platformy či operačného systému. Platformu MHP vyvíjajú experti DVB. Je určená predovšetkým pre digitálne vysielače využívajúce štandard DVB-T. Pre skutočné globálne využitie MHP bolo nutné pokryť i destináciu ako je Japonsko, USA, kde se využívaju iné štandardy digitálneho vysielania než DVB. Preto bola vyvinuta špecifikacia **GEM (Globally Executable MHP).** Táto špecifikácia vychádza z MHP, ale sú z nich odstránené špecifikácie štandardu DVB. Tým je umožnená kompatibilita i s inými štandardami[12].

D. Interaktívne služby v MHP

Hlavnou podstatou interaktívneho televízneho systému je schopnosť spustiť aplikácie, ktoré boli stiahnuté z vysielaného datového toku. Jednou zo základných myšlienok MHP je neexistencia samostatnej MHP aplikacie. To znamená, že každá MHP aplikacia je riadená nejakou službou. Kedže divák využívá MHP aplikáciu v spojení s určitým televizným kanálom a tento kanál prepne, dôjde väčšinou k jej vypnutiu. To vedie k obmedzeniu aplikacií, ktoré môžu bežať neustále, bez závislosti na určitom kanáli (službe). Poskytovateľom tak dokážu lepšie zaručiť, že aplikácie budú správne pracovať. Tento prístup zároveň umožňuje divákovi navigáciu medzi aplikáciami. Dostupné budú len tie, ktoré súvisia s aktuálne sledovaným programom (napr. hlasovanie). Iba vysielací súbor tvoriaci aplikáciu nepostačuje. MHP prijímač musí zaistiť k akej aplikácii dané súbory patria a ako sa má nimi zaobchádzať. MHP preto definuje informačnú aplikačnú tabuľu AIT. Tá prijímač informuje o aktuálne dostupných aplikáciách. Tieto informácie zhromažďuje komponenty a aplikačný manažer v MHP prijímači. Okrem sledovania dostupných služieb, je aplikačný manažér zodpovedný za monitorovanie aktualných služieb, ktoré môže aplikácie spustiť či ukončiť.

MHP aplikácie môžu byť napísané buď v Jave alebo HTML. Platforma MHP okrem toho definuje množstvo rozšírení a obmedzení v schopnostiach oboch jazykov. Z tohoto dôvodu nie sú aplikácie MHP zlučiteľné so štandardnou Javou či HTML. Pre odlíšenie sa teda MHP verzia Java aplikácie označuje DVB–J a HTML verzia DVB-HTML[13],[14].

E. HTML aplikácie

Pridaním HTML podpory do MHP bude umožnený prístup k internetovým stránkam. Väčšina HTML stránok je vytvorená pre rozlíšenie väčšie než 1024x768, ale rozlíšenie bežnej televízie je iba 720x576, čo spôsobí celkovú neprehľadnosť.

Application Infomation Table (AIT). Ako už bolo uvedené, AIT informuje o dostupných aplikáciách. To však nie je jediná informácia, ktorú poskytuje, ďalej su tu riadiace kódy: AUTOSTART, PRESENT, DESTROY, KILL, PREFETCH, REMOTE.

Aplikácia, ktorá obsahuje kód AUTOSTART, bude spustená automaticky po načítaní do prijímača. To umožňuje spustiť aplikácie, ktoré priamo súvisia s práve vysielaným programom.

Pokiaľ prijímač príjme aplikácie s riadiacim kódom PRESENT, uloží ju do zoznamu dostupných aplikácií a užívateľ si ich môže prípadne spustiť.

Aplikácie s riadiacim kódom KILL alebo DESTROY sú prijímačom vypnuté. To je možné využiť pri skončení určitého programu, z ktorým bola aplikácia zviazaná.

Rozdiel medzi kódom KILL a DESTROY je, že pri KILL má užívateľ možnosť v danej aplikácií pokračovať. Pri kóde DESTROY je aplikácia vždy uzatvorená.

Kód REMOTE prikazuje prijímaču prejsť na určitú službu za účelom spustenia aplikácie poskytovanej danou službou.

Dalšia dôležitá informácia udáva pre akú verziu MHP je aplikácia určená. To dáva prehrávaču možnosť rozhodnúť, kedy je daná aplikácia schopná správne interpretovať a teda či ju zaráta do zoznamu dostupných služeb alebo nie [15],[16].

III. VYSIELANIE MHP APLIKÁCIÍ

V dnešnej dobe je najčastejšou metódou vysielania aplikácií DSM-CC objektový karusel. **Digital Storage Media - Command and Control** (***DSM-CC***). resp. objektový karusel je transportný protokol (suborový systém), prostredníctvom ktorého je možné prenášať aplikácie medzi distribútorom a prijímačom. Tento prenos využíva zakódovanie aplikácií do hlavného vysielacieho toku spolu s video tokom (MPEG2), audio tok dátami (MPEG1 layer II) a riadiacimi dátami. Datový tok je zložený z dvoch základných častí. Prvá časť tvorí samotné data aplikací (jednotlivé soubory). Druhá časť prenáša informačnú tabuľku aplikácií AIT (Application Information Table), ktorá obsahuje informácie priraďujúce dané súbory konkrétnej aplikácií. AIT môže dalej obsahovať riadiace kódy: **AUTOSTART, PRESENT, DESTROY, KILL, PREFETCH, REMOTE**. Tieto kódy boli popísane vyššie.

Objektový karusel je určitý balík aplikácií vysielaný v neustálej slučke. Tento spôsob prenosu prináša problém s množstvom dát prenášených v karusele, toto množstvo je totiž priamo úmerné dĺžke periódy jednej slučky, behom ktorej sa zopakuje celý prenos dát. Dĺžka periódy určuje, akú najdlhšiu dobu bude musieť prijímač čakať, než bude môcť začať sťahovať aktuálne dáta. V praxi sa používa kompromis, ktorý vyčká prijateľný čas potrebný k načítaniu celého karuselu.

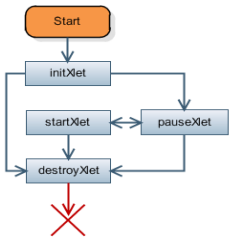
A. Java TV

Aplikácie určené pre MHP nejčastejšie programované v upravenej verzi jazyka Java, sú označované ako DVB-J a upravená Java nesie označenie JavaTV. Hlavným rozdielom oproti štandardnej Jave je implementácia balíka javax.tv a jeho verzii. Tu sa nachádza základna trieda radiaceho behu DVB-J aplikácií. Tieto aplikácie sú samozrejme vytvorené triedami, ktoré nie sú obsadene v javax.tv. Napríklad spätný kanál je programovany výhradne triedami z java.net.

B. Xlet

Pri aplikácii vytvorenej v štandardnej Jave sa predpokladá, že v reálnom čase bude prevedená len jediná aplikácia. Digitálna televízia však potrebuje spustiť viac aplikácií v rovnakom čase**.** Applet je spustený z webového prehladača a zavádza tzv. životný cyklus. Ten umožňuje aplikácie zastaviť (ak užívateľ prepne danú webovou stránku) a ukončiť ju (pokiaľ je webová stránka s appletom uzavretá). Zároveň môže být spustených niekoľko appletov súčasne. Applet bol preto prepracovaný pre podmienky digitálnej televízie a výsledok sa nazýva Xlet. Xlet teda obsahuje životný cyklus (Obr.3), ale oproti tomu je applet zložený zo stavu: „načitanie“, spušťanie“, pozastavenie“ a „ukončenie“.

Stav „pozastavenie“ je využitelný, pokiaľ je spustených viac aplikácií. V takom prípade bude viditeľná iba tá aplikácia, s ktorou práve pracujeme. Aplikácie na pozadí prejdu do režimu „pozastavenie“ a uvolia zdroje (pamäť) pre aktuálne výpočty [12].



Obr. 3 Životný cyklus Xletu.

IV. IMPLEMENTÁCIE MHP V PRAKTICKÝCH APLIKÁCIÁCH

A. Java TV rozhranie a možnosti použitia

Java TV je Java ME, na technológii, ktorá poskytuje výkonnejšie, bezpečné a ľahko implementovane riešenie pre vývoj Java aplikácií, ktoré bežia v televízii a nastavenia zariadenia set top box. Java Tv ponuka vývojárom možnosť vytvárať aplikácie,ako je elektronický programový sprievodca (EPG), Video-on-Demand (VOD) klientov, hry a výukové aplikácie, aplikácie pre prístup k internetu dáta (napr. počasie, správy, sociálnej siete), a na väčšine Blu-ray titulov, užívateľské rozhranie a bonusový obsah.

B. API References – programátorské jadro systému Java TV

javax.tv.graphics Pridáva podporu AWT pri otázkach špecifických pre televíziu, ako premiešavanie grafiky videa a objavovanie hlavného kontajnera s aplikáciou.

javax.tv.locator Poskytuje prostriedky pre odkazovanie na dáta a zdroje dostupné prostredníctvom Java API TV.

javax.tv.media Poskytuje ovládanie a akcie pre riadenie v reálnom čase médiá v televíznom prostredí

javax.tv.net Poskytuje mechanizmus prístupu k IP datagramom vo vysielanom dátovom toku.

javax.tv.service Poskytuje API, predstavuje SI prvkov

javax.tv.service.guide Poskytuje API pre podporu elektronického programového sprievodcu (EPG), vrátane programových plánov, programov podujatia a hodnotenia programu.

javax.tv.service.navigation Poskytuje API pre pohyb služieb a hierarchické servisné informácie.

javax.tv.service.selection Poskytuje mechanizmus pre prezentáciu výberu služieb.

javax.tv.service.transport Poskytuje ďalšie informácie o transportných mechanizmoch, ktoré poskytujú obsah SI údajov.

javax.tv.util Poskytuje rozhranie API pre vytváranie a správu časovača.

javax.tv.xlet Poskytuje rozhranie používané aplikáciami a Správcu aplikácií pre komunikáciu [14].

Praktické použitie Java TV

Umožňuje rôznu interaktivitu prístupu EPG, VIDEO –on –Demand, pripojenie na internet, sledovanie Facebook, Skype. V prijímačoch Samsung HDTV založených na App Store, umožňuje spustiť 3D aplikácie Video On Demand [17].

ZÁVER

V článku boli predstavené možnosti použitia IPTV technológie v dnešnej dobe. V príspevku bol predstavený prehľad možných služieb, ktoré poskytuje technológia IPTV. Zároveň je podrobnejšie opisaná MHP služba. Je charakterizovaná po stránke prenosu medzi prijímačom a vysielačom. Tento článok zároveň opisuje jednotlivé vrstvy technológie MHP a jej možnosti využitia v LCD TV. Taktiež oboznámuje s požiadavkami aké sú kladané na službu MHP. Potreba poznatkov je zrejmá z dôvodu veľkého rozšírenia týchto technológií v poslednej dobe. Aplikácia MHP je realizovaná s možnosťami v zariadeniach ako je napr. Set-box s MHP službami, a zároveň služby v LCD TV prijímačoch. Niektoré tieto technológie poskytujú ľahký a optimálny prístup k funkčnosti celého MHP rozhrania. Vizualizácia a funkčnosť implementácie MHP aplikácie je popísaná v HTML forme spolu s technológiou Java TV, jej rozhranie a možnosti použitia. Pre dnešnú dobu je potrebná veľká implementácia spomínaných technológií a to je možné len pomocou testovania na konkrétnych parametroch prenosu ako sú prenosová rýchlosť, spoľahlivosť, straty a výpadky prenosu. Je potrebné sa zaoberať týmito parametrami a teda je na mieste skúmať spoľahlivosť prenosu a ďalšie potrebné znalosti a parametre efektívnej služby MHP.

LITERATÚRA

[1] The ITV dictionary : http://www.itvdictionary.com/itv.html, 2001-2003
[2] Ben, “Co bychom mohli mít a o co zatím přicházíme”, Sděl. technika 10/2004, p. II
[3] E. Schwalb,” iTV Food Chain”, Pearson Education, Prentice Hall PTR, Indiana.
[4] M. Legiň,” Televizní technika DVB-T” Praha, technická literatúra BEN, 2006. ISBN 80-7300-204-3
[5] J. Lipták.**“**Systémy digitálneho televízneho vysielania”, Masarikova univerzita, Fakulta informatiky, 2010
[6] www.digizone.cz
[7] P. Chmiel, T. Kostka, E.Navrátilová, “Televizní Přenosový řetezec”, Havířov-Šumbark
[8] P. Kasal, ”Interaktivní aplikace pro DVB-MHP”, Vysoke učeni technické v Brne, Fakulta elektrotechniky a komunikačních technologií, 2008/2009
[9] S.Morris, A.Chaigneau, ”Interactive TV standards,Amsterdam,” Elsevier, ISBN 0-240-80666-2, s. 585, 2005
[10] M. Legiň,” Televizní technika DVB-T”, Praha: BEN, ISBN 80-7300-204-3, s. 286, 2006
[11] F. Šot, “MHP application for remote control of actuating unit”, BRNO University of technology, 2008
[12] V. Švanda, “Methods for image enhancement in interactive applications of digital telecasting”, BRNO University of technology, 2008
[13] J.-P. Evain, "The Multimedia Home Platform” www.dvb.org\documents\white-papers\evain.pdf , 2006
[14] S. Morris,“ The JavaTV”,Tutorial”, www.interactivetvweb.org/tutorials/ javatv/javatv_intro
[15] S. Morris, A. Chaigneau, ” Interactive TV standards” , ISBN 0-240-80666-2, 2005
[16] S. Morris “The MHP Tutorial” , www.interactivetvweb.org/ tutorial/mhp/index.shtml.htm, 2007
[17] www.engadget.com/2010/08/11/samsung-puts-up-cash-hdtvs-for-developers-in-free-the-tv-apps

Vyhladzovanie textúr 3R modelu ľudskej hlavy

Ondrej KOVÁČ, Ján MIHALÍK

Katedra elektroniky a multimediálnych telekomunikácií, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

ondrej.kovac@tuke.sk, jan.mihalik@tuke.sk

***Abstrakt* — V článku sú popísané základné princípy využité pri generovaní textúry 3R modelu ľudskej hlavy z dvoch ortogonálnych pohľadov na ľudskú hlavu. Taktiež je popísaná základná kubická transformácia. Hlavným predmetom záujmu tohto článku je problematika odstránenia nežiaducich rušivých vplyvov hraníc prechodov, ktoré sú priamym dôsledkom spájania dvoch ortogonálnych pohľadov.**

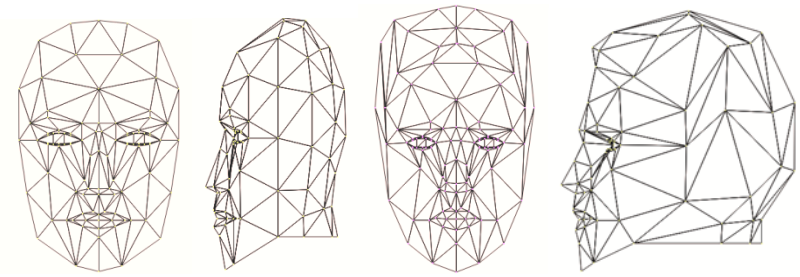
***Kľúčové slová* — 3R model, textúra, klon, vyhladzovanie, transformácia.**

I. ÚVOD

V dobe, keď sa čoraz väčší dôraz kladie na sociálne siete slúžiace na komunikáciu medzi ľuďmi a virtuálne reality slúžiace k zábave v rôznych on-line hrách, je nevyhnutné vyriešiť problematiku postáv reprezentujúcich reálnych ľudí. Tieto postavy sa nazývajú klony. Sú to virtuálne repliky človeka sediaceho za počítačom. V tomto článku sa budeme zaoberať iba problematikou ľudskej hlavy. K vytvoreniu takejto hlavy klonu je potrebný základný model ľudskej hlavy, ktorý sa upraví na základe anatomických čŕt daného človeka a textúru, ktorá reprezentuje výzor človeka. Takúto textúru získame z dvoch ortogonálnych snímok ľudskej hlavy. Pretože vytvárame textúru z viacerých snímok, vznikajú na hrane ich spojenia prechody, ktoré pôsobia rušivo a preto ich treba vyhladiť.

II. TEXTÚRY 3R MODELU ĽUDSKEJ HLAVY

Textúrovanie v problematike kódovania obrazu ľudskej hlavy na báze modelovania, dáva jej tvarovanému modelu konečnú a prirodzenú podobu. [1][2][3][4][10] Textúra teda predstavuje jasové úrovne obrazových prvkov (op), ktoré sú výsledkom spojenia dvoch a viac referenčných snímok. Najčastejšie ide o čelný a profilový pohľad. Textúrou teda dosahujeme vyššiu mieru podobnosti vygenerovaného klonu a jeho predlohy. Samotné textúrovanie 3R modelu ľudskej hlavy je teda procesom prenášania textúrovej hodnoty op na už prispôsobený model. Výhodou textúry v modelovom kódovaní je to, že textúra sa nemení tak, ako sa nemení vzhľad človeka, ale menia sa iba podmienky prostredia. Z toho vyplýva, že v procese videosekvencie postačuje, ak je textúra vyslaná na prijímaciu stranu iba raz. Keďže anatómia ľudskej hlavy je u každého človeka odlišná, je potrebné mať k dispozícii rôzne modely, ktoré sú prispôsobené na človeka, ktorého klon tvoríme. Takéto prispôsobenie nazývame tvarovanie. Pohľad na model pred tvarovaním je na obr.2.1 a pohľad na model po tvarovaní na obr.2.2.



Obr.2.1 Čelný a profilový pohľad na štandardný 3R model Obr.2.2 Čelný a profilový pohľad na tvarovaný 3R model

Každý model je definovaný sadou vertexov a ich vzájomným prepojením. Na aproximovanie povrchu uzatvorených triangulácii je potrebná vhodná transformácia súradníc, aby sa neovplyvnili príspevky polygónov, ktoré sú po priemete do hv (v anglosaskej literatúre uv) roviny vzájomne prekryté, ale nie sú susedné. Transformáciu je možné vykonať pomocou otvorenej transformácie, pri ktorej je pred samotnou transformáciou nutné model rozpojiť v jeho zadnej časti. Taktiež je možné využiť uzatvorenú transformáciu, pri ktorej sa model transformuje bez predošlej úpravy. Uzatvorenou transformáciou sa ďalej nebudeme zaoberať.

III. KUBICKÁ TRANSFORMÁCIA

Kubická transformácia súradníc (h,v,r) → (h_N ,v_N) je definovaná pomocou vzťahov 3.1-3.2. [6]

$$h_N = \begin{cases} h + m_1 \left[m_2 \left(1 - \frac{r - r_{min}}{|r_{max} - r_{min}\square|} \right) \right]^3 & h > 0 \\ h - m_1 \left[m_2 \left(1 - \frac{r - r_{min}}{|r_{max} - r_{min}\square|} \right) \right]^3 & h < 0 \end{cases} \quad (3.1)$$

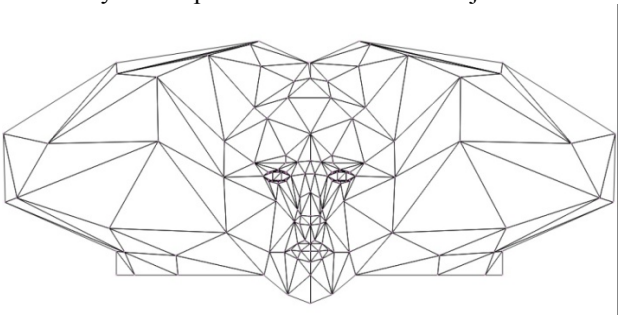
$$v_N = m_3 v \quad (3.2)$$

kde konštanty m_1, m_2, m_3 slúžia na nastavenie parametrov transformácie, r_{min} , r_{max} sú minimálna a maximálna r súradnica v triangulácii.
Zo vzťahov je zrejmé, že funkcia tretej mocniny po ktorej táto transformácia nesie meno je aplikovaná len na horizontálnu súradnicu. Vertikálna súradnica je lineárne zväčšená konštantou m_3. Ak by sme vertikálne súradnice nechali nezmenené, bolo by možné túto transformáciu využiť, ale vplyvom prílišnej deformácie by dochádzalo ku skresleniam textúry. Ďalej si treba uvedomiť, že transformáciu nie je nutné vykonávať pre všetky vertexy. [7] Oblasť tváre je vhodné z transformácie vynechať a iba ju lineárne zväčšiť konštantou m_3. Takže pre oblasť tváre budú platiť vzťahy 3.3, 3.4 a pre ostatné oblasti vzťahy 3.1 a 3.2.

$$h_N = m_3 h \quad (3.3)$$

$$v_N = m_3 v \quad (3.4)$$

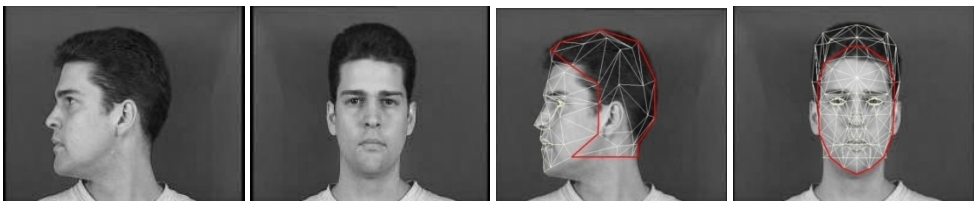
Transformovaný tvarovaný model potom bude mať tvar ako je zobrazené na obr.3.1.



Obr.3.1 Tvarovaný 3R model transformovaný kubickou transformáciou

IV. GENEROVANIE TEXTÚR 3R MODELU ĽUDSKEJ HLAVY POMOCOU DVOCH ORTOGONÁLNYCH POHĽADOV.

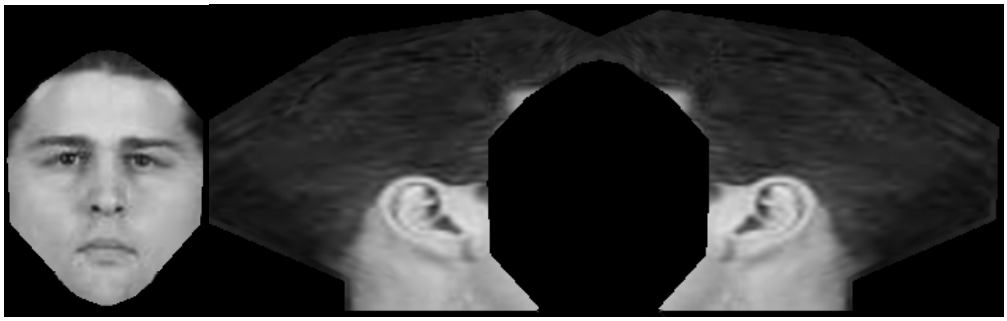
Vzhľadom na to, že ľudská hlava je za symetrická, na vytvorenie klonu sú postačujúce dve snímky a to snímka čelného a jedného profilového pohľadu na ľudskú hlavu. V prípade, ak sú k dispozícii iba dva pohľady, je nevyhnutné, aby tieto boli ortogonálne. Tieto pohľady sú zobrazené na obr.4.1.a,b. Snímky sú následne preložené priemetom tvarovaného modelu a niektorou zo známych metód sú obrazové prvky pod jednotlivými polygónmi prenesené na príslušné polygóny v transformovaného modelu. na obr. 4.1.c,d



a) b) c) d)

Obr.4.1 Snímka a) profilového pohľadu b) čelného pohľadu
d) profilového pohľadu preložená modelom b) čelného pohľadu preložená modelom

Je zrejmé, že profilový pohľad obsahuje aj časť čelného pohľadu a naopak. Preto je nutné určiť relevantnú časť profilového a čelného pohľadu. Tak ako je znázornené na obr.4.1.c,d. Ak sa generuje textúra na základe takto striktne definovaných pravidiel, výsledná textúra predstavuje spojenie dvoch čiastkových textúr obr.4.2. Takáto textúra obsahuje viditeľné poruchy na prechode z čelného do profilového pohľadu. Tieto poruchy sú dôsledkom rozličných jasových a detailových vlastností oboch snímok. Takáto textúra je znázornená na obr.4.3. Na obr.4.4 je výsledný model textúrovaný takouto textúrou. [10]



Obr.4.2 Čiastkové textúry čelného a profilového pohľadu



Obr.4.3 Výsledná textúra s viditeľným skreslením



Obr.4.4 Klon s nevyhladenou textúrou.

V. VYHLADZOVANIE TEXTÚR 3R MODELU ĽUDSKEJ HLAVY

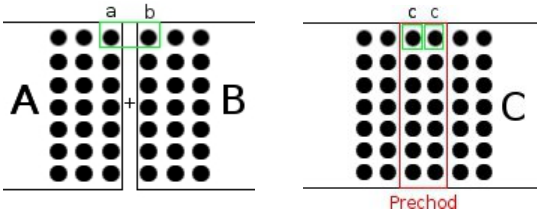
Existuje mnoho techník pre vyhladzovanie obrazu rsp. spájanie viacerých obrazov tak, aby hranica prechodu bola úplne odstránená, alebo aspoň potlačená na minimálnu možnú mieru. [5] [9] Pred samotnou aplikáciou ktoréhokoľvek algoritmu vyhladenia hrán je žiaduce vstupné obrazy v rámci jasového rozsahu čo najviac vzájomne prispôsobiť. Takúto úpravu obrazu dosiahneme využitím centrovania a energetického normovania. [8] Toto je popísané vzťahom 5.1. Kde **X** predstavuje raster textúry po úplnom predspracovaní.

$$\underline{\mathbf{X}} = \frac{\underline{\mathbf{Y}} - \alpha}{\sigma} \qquad \alpha = \frac{1}{n}\sum_{i=1}^{n} Y_i \qquad \sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \alpha\right)^2} \tag{5.1}$$

kde $\underline{\mathbf{Y}}$ sú hodnoty necentrovanej textúry bez energetického normovania. Konštanta α je stredná hodnota a σ smerodajná odchýlka textúry.

A. Vyhladzovanie metódou váhovania

Ide o najjednoduchšiu techniku vyhladzovania pri ktorej sa hraničné obrazové prvky (rušivé prvky) výsledného obrazu nahradia hodnotou získanou z posledných nenulových prvkov spájaných obrazov pomocou vzťahu 5.2, tak ako je zobrazené na obr.5.1.



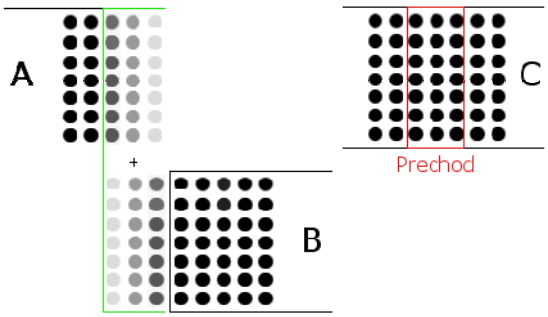
obr.5.1 Princíp vyhladzovania metódou váhovania

$$c = \frac{a + b}{k} \tag{5.2}$$

Kde ***c*** je hodnota prvkov tvoriacich prechod v danom riadku, ***a, b*** sú hodnoty op spájaných obrazov A a B, ***k*** je reálna konštanta z intervalu <1;2>.
Konštanta ***k*** zabezpečuje prispôsobenie parametrov pre výpočet hodnoty prvkov prechodu pre každý riadok. Je teda zrejmé, že konštanta ***k*** sa bude meniť vzhľadom k jasovým zmenám v riadkoch.

B. Vyhladzovanie metódou prekrývania

Pri tejto technike sa vstupné obrazy upravujú nasledovne: Obrazu A sa v cieľovej oblasti prechodu lineárne zmenšuje hodnota jasu obrazových prvkov. Keďže obrazu A sa nezmenila veľkosť rastra a má dôjsť k prekrytiu oboch obrazov, je potrebné obrazu B doplniť obrazové prvky. Týchto prvkov bude práve toľko, koľko je potrebné na oblasť prechodu. Dopĺňame hodnotu posledného obrazového prvku obrazu B ktorej hodnoty postupne lineárne znižujeme. Toto je ilustrované na obr .5.2.



Obr.5.2 Princíp vyhladzovania metódou prekrývania

VI. EXPERIMENTÁLNE VÝSLEDKY

Vstupné obrazy obr. 4.2 boli predspracované pomocou centrovania, výkonového normovania a následne na to spojene. Na nasledujúcom obraze je znázornená výsledná textúra po centrovaní a normovaní vstupných čiastkových textúr.



Obr.6.1 Predspracovaná textúra

Z obr. 6.1 je zrejmé, že po procese normovania dochádza v oblasti líc k čiastočnému vyhladeniu prechodu. Na takto predspracované vstupné čiastkové textúry boli aplikované vyššie popísané algoritmy vyhladenia. Textúra vyhladená pomocou metódy váhovania je na obr. 6.2.



Obr.6.2 Textúra vyhladená metódou váhovania

Ako je z obr. 6.2 vidieť, pri tejto metóde dochádza stále ku skresleniu v oblasti kútov. Ďalej sme experimentom zistili, že optimálne okno prechodu je tvorené dvoma op. Na obr. 6.3 je zobrazený výsledný klon textúrovaný textúrou vyhladenou metódou váhovania.



Obr.6.3 Klon s textúrou vyhladenou metódou váhovania.

Na nasledujúcom obrázku je zobrazená textúra vyhladená metódou prekrývania. Oblasť prechodu pozostáva zo 4 op.



Obr.6.4 Textúra vyhladená metódou prekrývania

Na obr. 6.5 je zobrazený výsledný klon textúrovaný textúrou vyhladenou metódou prekrývania.



Obr.6.5 Klon s textúrou vyhladenou metódou prekrývania

VII. ZÁVER

Využitím popisovaných algoritmov sme dosiahli istú mieru potlačenia rušivého vplyvu hrán na celkovú subjektívnu kvalitu výsledného klonu. Žiaľ k úplnému odstráneniu hrán nedošlo. Vyššia subjektívna kvalita klonu bola dosiahnutá aplikáciou metódy vyhladzovania pomocou čiastočného prekrývania vstupných obrazov. Preto usudzujeme, že vhodnou kombináciou nastavení šírky prechodu a prípadným zavedením nelinarity do znižovania (zvyšovania) jasu op spájaných obrazov v oblasti prechodu sa dajú dosiahnuť lepšie výsledky. Využitie centrovania a energetického normovania sa ukazuje ako žiaduce predspracovanie spájaných obrazov, pretože eliminuje vplyvy svetelných podmienok pri získavaní referenčných snímkou. pDosiahnuté výsledky je možné vylepšiť aj aplikáciou dvojrozmerného Gaussovho filtra vyššie popísanými algoritmami na vyhladenú textúru. Ten je potrebné aplikovať iba na oblasť prechodu a jeho najbližšie okolie, aby sa zamedzilo skresleniu aj tých častí, ktoré už nie sú súčasťou prechodu.

LITERATÚRA

[1] THALMANN, N. M., THALMANN, D. Handbook of Virtual Humans. Chichester (England): Wiley, 2004.

[2] Mihalík,J.-Michalčin,V.: 3D Motion Estimation and Texturing of Human Head Model. Radioengineering, Vol.13, No.1, 2004, p.26- 3.

[3] Mihalík,J.-Michalčin,V.: Texturing of Surface of 3D Human Head Model. Radioengineering, Vol.13, No.4, 2004, p.44- 47.

[4] Yong-Jin Liu, Matthew Ming-Fai Yuen.: Geometry-optimized virtual human head and its applications. Computers & Graphics, Volume 32, Issue 6, December 2008, Pages 624-631

[5] Burt, J.P. Adelson, E.H. Amultiresolution spline with application to image mosaics, ACM Transactions on Graphics, Vol. 2, No. 4, October 1983, Pages 217-236.
http://graphics.cs.cmu.edu/courses/15-463/2005_fall/www/Papers/BurtAdelson.pdf

[6] Kováč O.:Generovanie a kódovanie textúr ľudskej hlavy. Diplomová práca FEI TU Košice, 2011.

[7] Lee W.S.-Thalmann M.: Head Modeling from Pictures and Morphing in 3D with Image Metamorphosis based on triangulation.
http://www.site.uottawa.ca/~wslee/publication/CAPTECH98.pdf

[8] Mihalik,J.-Kasár,M.: Basis of Eigenfaces for Tracking of Human Head. Journal of Electrical Engineering, Vol.58, No.3, 2007, p.134-139.

[9] PRAT, W. K. Digital Image Processing. John Wiley & Sons. New York, 1978.

[10] http://www.blender.org/

Kombinovaná digitálna vodotlač v statických obrazoch

Zita KLENOVIČOVÁ, Ján CHALUPKA, Dušan LEVICKÝ

Katedra elektroniky a multimediálnych telekomunikácií, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

zita.klenovicova@tuke.sk, jan.chalupka@student.tuke.sk, dusan.levicky@tuke.sk

***Abstrakt* — Predkladaný článok sa zaoberá kombinovanou digitálnou vodotlačou v statických obrazoch, ktorá využíva vkladanie jednej časti vodoznaku do obrazového priestoru a druhej časti vodoznaku do transfornovaného priestoru . Dosahuje sa tým zvačšenie kapacity vloženej informácie a vložený vodoznak má vlastnosti krehkých vodoznakov, ktoré sa vkladajú do obrazového priestoru aj vlastnosti robustných vodoznakov, ktoré sa vkladajú do transformovaného priestoru.**

***Kľúčové slová* — digitálna vodotlač,obrazový priestor, transformovaný priestor, vkladanie a extrakcia vodoznaku, DCT transformácia**

I. ÚVOD

Dlho predtým, ako sa vodoznak začal používať v súvislosti s digitálnymi obrazmi, sa pod pojmom vodoznak obvykle rozumel obrazec, či vzor na papieri, ktorý sa pri nasvietení prejavoval ako rôzna kombinácia svetlých a tmavých miest. To bolo spôsobené typicky rôznymi hrubými papiermi. Už papierové vodoznaky zdieľajú s digitálnymi vodoznakmi jednu dôležitú vec – niektoré sú zreteľné hneď, iné až po bližšom pohľade a iné si pre zmenu bez nasvietenia nevšimnete vôbec. Vodoznaky sa v papierovej podobe dodnes používajú v bankovkách, pasoch, známkach a ďalších dôležitých dokumentoch ako jedna z prevencií proti falšovaniu. Skúsená osoba je totiž schopná pomerne spoľahlivo rozoznať originál a falzifikát.

Na rozdiel od papierových vodoznakov, digitálny vodoznak (digital watermark) je identifikačný kód, ktorý je natrvalo vložený do digitálnych dát a nesie informáciu týkajúcu sa ochrany autorských práv pri kopírovaní dát a pri autentifikácie dát. Takže aj digitálne vodoznaky, tak ako aj papierové vodoznaky, vznikli ako prevencia proti falšovaniu. Digitálna vodotlač (digital watermarking) umožňuje identifikovať zdroj, autora, tvorcu, vlastníka, distribútora alebo autorizovaného spotrebiteľa digitálnych obrazov, video záznamov alebo audio záznamov.

II. DIGITÁLNA VODOTLAČ

Digitálna vodotlač možno definovať ako techniku ukrytia prídavnej informácie do rôznych druhov modalít multimédií. Prídavná informácia (vodoznak) može reprezentovať údaje o vlastníkovi autorských práv, resp. o distribútorovi označenej modality a taktiež slúži ako prostriedok na zabezpečenie integrity multimédiálnych dát [1], [11],[2].

Vkladaný vodoznak môže mať rôznu formu a podobu v závislosti od použitia v aplikácii. Môže mať formu konečnej postupnosti symbolov, resp. číslic, prídavnej obrazovej informácie (loga), segmentu rečového signálu, poprípade môže mať formu 1-bitovej informácie, ktorá indikuje, či originálne dáta sú alebo nie sú označené vodoznakom.

Digitálna vodotlač musí spĺňať tri základné požiadavky, a to [6],[5]:

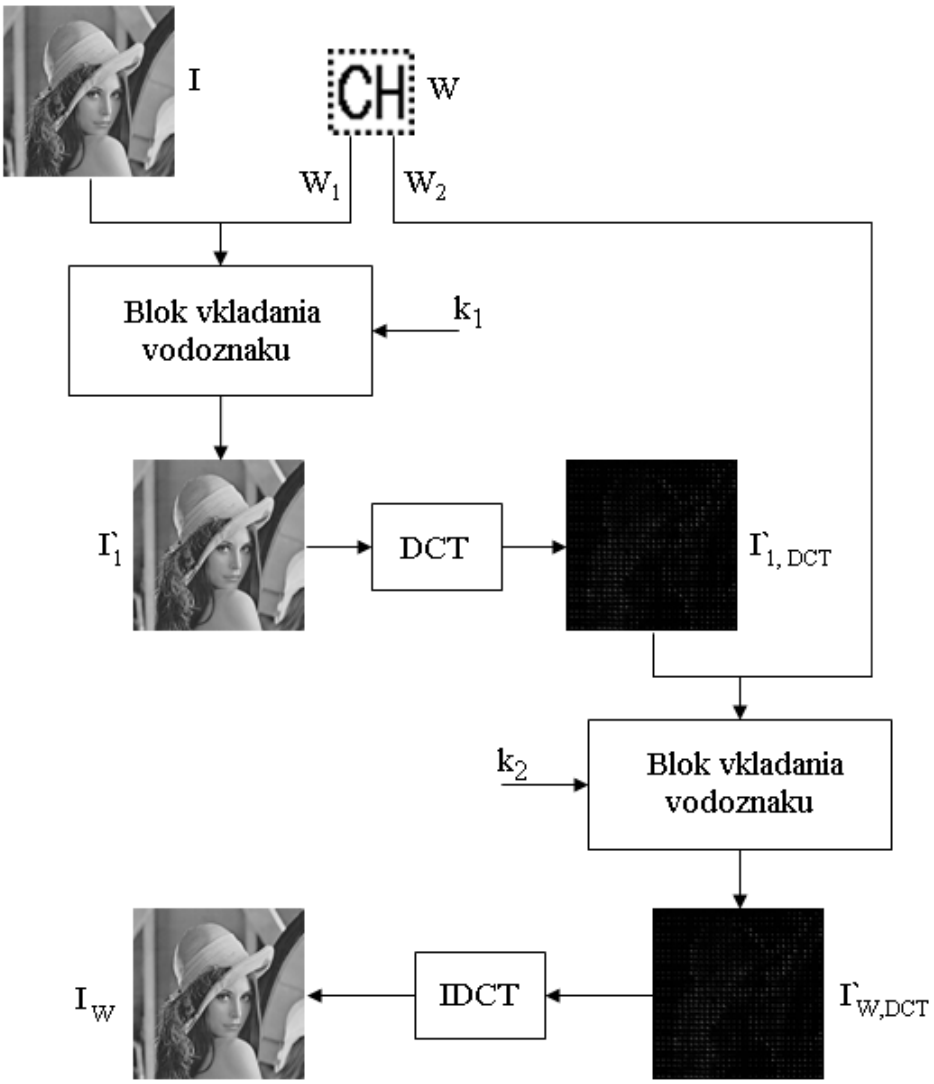
- robustnosť vodoznaku,
- perceptívna transparentnosť označených dát,
- množstvo vloženej informácie .

Pri návrhu metódy digitálnej vodotlače je nutné voliť vyhovujúci kompromis medzi týmito tromi požiadavkami.

Na realizáciu digitálnej vodotlače v statických obrazoch možno využiť v zásade obrazový, transformovaný a parametrický priestor [9],[7].

III. KOMBINOVANÁ DIGITÁLNA VODOTLAČ

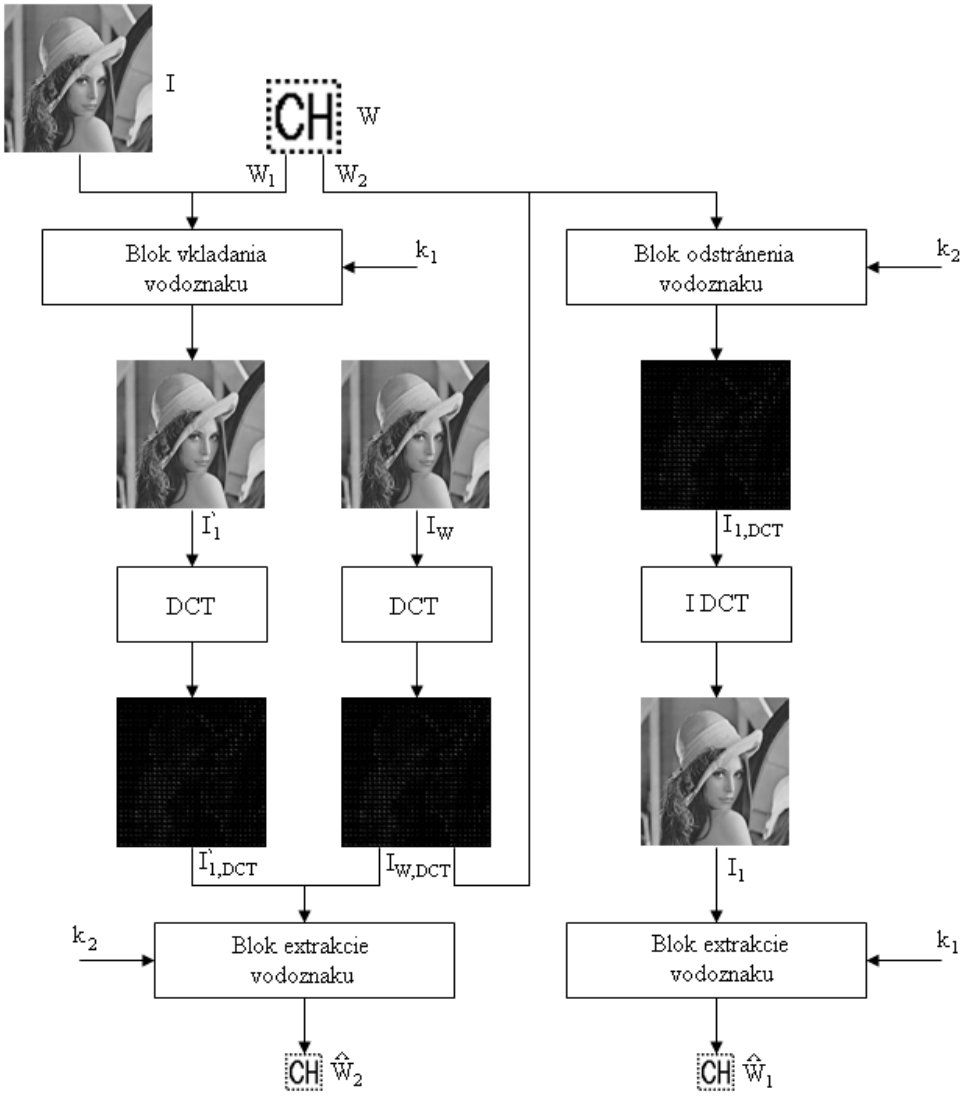
Kombinovaná digitálna vodotlač je založená na kombinácii vkladania vodoznaku do obrazového a do transformovaného priestoru [12].Ako transformovaný priestor bol zvolený priestor DCT. Jednoduchý model vkladania vodoznaku kombinovanou metódou vkladania vodoznaku je zobrazený na obr. 1.



Obr. 1 Schéma kombinovanej metódy vkladania vodoznaku

Vstupom vkladania vodoznaku je originálny obraz I o rozmere N1xN2 a vodoznak W o rozmere M1xM2. Vodoznak W sa rozdelí na 2 časti: W1 a W2. Prvá časť vodoznaku W1 sa vkladá do originálneho obrazu I metódami vkladania do obrazového priestoru (napr. vkladanie vodoznaku do bitových rovín) za pomoci kľúča k1, čím vznikne označený obraz I`1. Na označený obraz sa aplikuje transformácia (napr. DCT), čím sa získa obraz v transformovanom priestore I`1,DCT, do ktorého sa vhodnou metódou vloží druhá časť vodoznaku W2 za pomoci kľúča k2. Vznikne označený obraz I`W,DCT v transformovanom priestore, z ktorého po aplikovaní inverznej transformácie vznikne výsledný označený obraz IW.

Všeobecná schéma extrakcie vodoznakov z označených obrazov kombinovanou metódou vkladania vodoznakov je zobrazená na obr. 2.



Obr. 1 Extrakcia vodoznakov z označených obrazov kombinovanou metódou vkladania vodoznakov

Na extrakciu vodoznakov je potrebný originálny obraz I, vodoznak W (ktorý rozdelíme na W1 a W2), označený obraz IW a tajné kľúče k1 a k2. Začiatkom procesu extrakcie je vkladanie časti originálneho vodoznaku W1 do originálneho obrazu I za pomoci tajného kľúča k1 do obrazového priestoru. Získa sa tým označený obraz I`1, ktorý je potrebný na extrakciu vodoznaku z transformovaného priestoru. Takto označený obraz sa rozdelí na neprekrývajúce sa bloky o rozmere 8x8 a následne sa na každý blok aplikuje DCT transformácia, čím sa získa obraz I`1,DCT v transformovanom priestore. Označený obraz IW sa taktiež rozdelí na neprekrývajúce sa bloky o rozmere 8x8 a na každý blok sa aplikuje DCT transformácia. Získa sa označený obraz v transformovanom priestore IW,DCT. Tieto dva obrazy vstupujú do bloku extrakcie vodoznaku, kde sa za pomoci tajného kľúča k2 (definuje pozíciu vloženého vodoznaku) vypočíta rozdiel medzi obrazom I`1,DCT a IW,DCT a získa sa vodoznak W2.

Pre extrakciu vodoznaku $\hat{W}1$ z obrazového priestoru je potrebné odstrániť vodoznak, prípadne vodoznaky z transformovaného priestoru označeného obrazu IW. Toto odstráneniu sa vykonáva v bloku odstránenia vodoznaku, kde vstupom je označený obraz v transformovanom priestore IW,DCT, druhá časť originálneho obrazu W2 a tajný kľúč k2 , ktorý obsahuje informáciu o pozícii vodoznaku, prípadne vodoznakov v transformovanom priestore . Získa sa tým obraz v transformovanom priestore I1,DCT. Na tento obraz sa aplikuje inverzná DCT transformácia, čím sa získa obraz I1, ktorý vstupuje spolu s tajným kľúčom k1 (obsahuje informáciu o pozícii vodoznaku, prípadne vodoznakov v obrazovom priestore) do bloku extrakcie vodoznaku. Výsledkom je extrahovaný vodoznak $\hat{W}1$ z obrazového priestoru .

IV. ZÁVER

Kombinovaná digitálna vodotlač je výhodná tým, že sa do obrazu vkladá väčšie množstvo informácie. Vodoznaky vkladané do obrazového priestoru zaraďujeme do kategórie krehkých vodoznakov, preto nie je potrebné ich do obrazového priestoru vkladať veľa. Naopak vodoznaky vkladané do transformovaného priestoru sú veľmi odolné voči mnohým typom útokov, preto je výhodné vkladať do transformovaného priestoru väčšie množstvo informácie.

V. POĎAKOVANIE

Táto práca bola vytvorená realizáciou projektu Rozvoj Centra informačných a komunikačných technológií pre znalostné systémy (kód ITMS projektu: 26220120030) na základe podpory operačného programu Výskum a vývoj financovaného z Európskeho fondu regionálneho rozvoja.

LITERATÚRA

[1] LEVICKÝ, Dušan – HOVANČÁK, Rastislav – KLENOVIČOVÁ, Zita: Technika digitálnych vodoznakov. Princípy, systémy a použitie. Slaboprúdový obzor, Roč. 59, č. 1, 2002 s 1-5
[2] CHALUPKA, Ján: Digitálne vodoznaky v statických obrazoch a textoch, Bakalárska práca, TU FEI, Košice, 2009
[3] Gerhard C Langelaar et all: Watermarking Digital Image and Video Data. IEEE Signal Processing Magazine, September 2000, Vol. 17, No.5. pp. 20.-45.
[4] LEVICKÝ, Dušan: Multimediálne telekomunikácie, Elfa, Košice, 2002
[5] VALENČÍK, Ján: Ukrývanie dát v statických obrazoch, Diplomová práca, TU FEI, Košice, 2008
[6] CHIOU-TING, Hsu – JA-LING, Wu: Hidden Digital Watermarks in Images. IEE Transactions on Image Processing, Vol. 8, No. 1. January 1999, pp. 58-68
[7] PROKOPOVIČ, Jozef: Hierarchické metódy vkladania vodoznakov do statických viacúrovňových obrazov, Diplomová práca, TU FEI, Košice, 2004
[8] POLEC, Jaroslav a kol.: Vybrané metódy kompresie dát, Vydavateľstvo FABER, Bratislava, 1996
[9] PODSTRELENEC, Milan: Vodoznaky v statických obrazoch s využitím CDMA, Diplomová práca, TU FEI, Košice, 2002
[10] LEVICKÝ, Dušan: Telemetrické systémy, Elfa, Košice, 1993
[11] Borko Furt et all: Multimedia Encryption and Watermarking, Springer, 2005
[12] Chalupka,J: Kombinovan8 digit8lna vodotlač v statických obrazoch, KEMT, 2011

Návrh I-Q demodulátora pre UWB senzorový systém na báze M-postupností

Martin LIPTAJ, Matej ŽIGA, Pavol GALAJDA

Katedra elektroniky a multimediálnych telekomunikácií, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

martin.liptaj@tuke.sk, matej.ziga@student.tuke.sk, pavol.galajda@tuke.sk

Abstrakt **– Článok sa zaoberá návrhom I-Q demodulátora určeného ako rozširujúci modul pre UWB senzorový systém, založený na pseudonáhodnej M - postupnosti. I-Q demodulátor je navrhnutý tak, aby pracoval vo frekvenčnom pásme od 6 GHz do 8 GHz, čo je v súlade s frekvenčným pásmom vyhradeným pre UWB aplikácie, podľa európskej organizácie ECC a bol zároveň jednoducho rekonfigurovateľný pre frekvenčné pásmo určene pre UWB aplikácie v súlade s organizáciou FCC.**

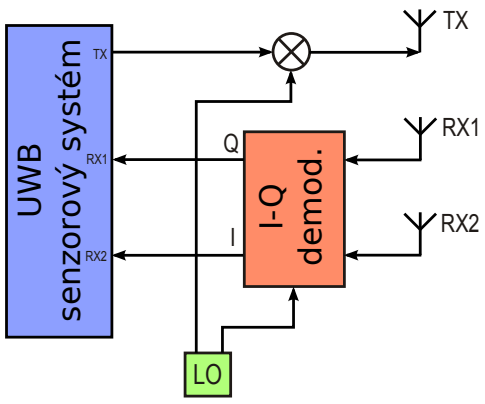
Kľúčové slová **– ECC, I-Q demodulátor, SiGe BiCMOS, UltraWideband, Zmiešavač**

I. Úvod

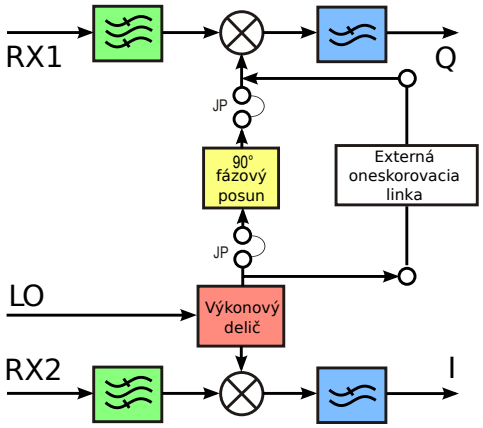
Jedným z významných míľnikov širokopásmovej (UltraWideBand-UWB) techniky bol rok 2002, kedy bolo inštitúciou Federal Communications Commission (FCC) umožnené s obmedzeným výkonom komerčne prevádzkovať UWB systémy vo frekvenčnom pásme od 3,1 GHz do 10,6 GHz. Vysielané spektrum UWB signálu je definované ako frekvenčné pásmo, ktoré je širšie ako 500 MHz, alebo viac ako 25% strednej resp. centrálnej frekvencie [1]. V tom období boli definované aj frekvenčné UWB pásma pre Európu organizáciou Electronic Communication Committee (ECC), ktorá takto uvoľnila frekvenčné pásmo od 6 GHz do 8,5 GHz [1].

Tieto UWB systémy nájdu uplatnenie v širokom spektre senzorovo orientovaných aplikácií. Je ich možné využívať v medicíne (detekcia rakovinových buniek v tele), v priemysle (zmena zloženia kvapalných a plynných zmesi), v bezpečnostných zložkách (lokalizácia osôb v budovách, hľadanie mín) a ďalších. Princípov činností UWB senzorových systémov je niekoľko. Jedným z možných riešení je UWB senzorový systém založený na pseudonáhodnej M - postupností [2].

UWB senzorový systém môže pracovať v základnom pásme (napríklad od niekoľkých MHz po jednotky GHz), ale jednoduchým spôsobom je možné systém rozšíriť a získať tak senzorový systém pracujú v preloženom pásme [3]. Rozšírenie môže byť vytvorené jednoduchým zmiešavačom na vysielacej strane a blokom I-Q demodulátora na strane prijímacej (obrázok 1). Návrh I-Q demodulátora prezentovaný v tomto článku je určený pre frekvenčné pásmo od 6 GHz do 8 GHz, čo je v súlade s ECC frekvenčným pásmom. Navyše, jednoduchou úpravou je možná aj prevádzka v FCC frekvenčnom pásme od 3,5 do 10,5 GHz.



Obr. 1 Zapojenie senzorového systému s rozširujúcim blokom



Obr. 2 Bloková schéma I-Q demodulátora

II. Základná teória I – Q demodulátora

Princíp I-Q demodulátora spočíva v rozdelení vstupného RF signálu a zmiešaní v dvoch zmiešavačoch s rôzným signálom lokálneho oscilátora (LO), z preloženého pásma do dvoch I a Q signálov v základnom pásme. Tieto I a Q signály obsahujú informácie o amplitúde a fáze dekódovaného vstupného signálu. Signály LO jednotlivých zmiešavačov sú s rovnakou frekvenciou, ale vzájomne fázovo posunuté o 90°. Na vstupe demodulátora sa spravidla nachádzajú pásmové priepusty, ktoré vymedzia požadované vstupné frekvenčné pásmo. Na výstupe I-Q demodulátora sa nachádzajú dolno-priepustné filtre, ktorých úlohou býva eliminácia rušivých produktov po zmiešaní, ako aj zložky LO signálu na výstupe zmiešavačov. I a Q signály bývajú spravidla digitalizované v A/D prevodníkoch a následne spracované v obvodoch číslicového spracovania signálov. Nevýhod analógového riešenia I-Q demodulátora je viacero. Keďže ide o paralelnu štruktúru, aj minimálne rozdiely medzi jednotlivými časťami vytvoria podstatnú chybu v meraniach. Tiež fázový posun medzi signálmi LO na jednotlivých portoch zmiešavačov musí byť presne 90°. To kladie veľký dôraz na kvalitu oneskorovacej linky. Tieto nevýhody sa dajú odstrániť použitím digitálneho I-Q demodulátora. Napriek tomu, tento článok sa zaoberá návrhom a výrobou analógového I-Q demodulátora.

III. Návrh I-Q demodulátora

Na obrázku 2 je možné vidieť blokovú schému I-Q demodulátora. Je zložený z dvojice zmiešavačov, oneskorovacej 90° linky, širokopásmového odporového deliča výkonu, vstupných a výstupných filtrov.

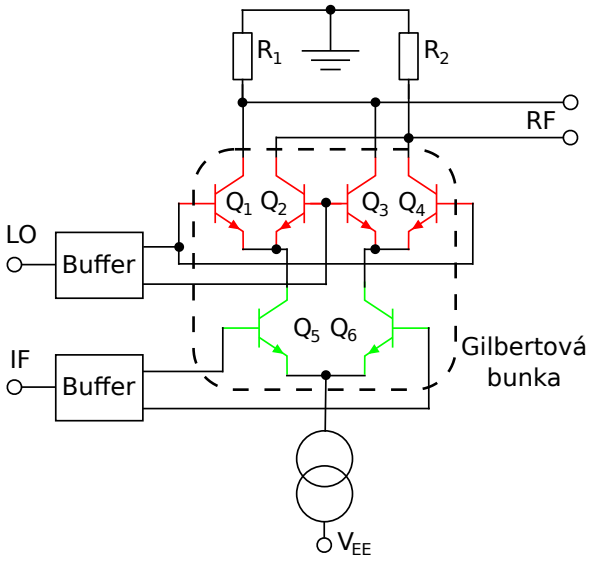
A. Zmiešavač

Pre potreby I-Q demodulátora na Katedre elektroniky a multimediálnych telekomunikácií, Fakulty elektrotechniky a informatiky, Technickej univerzity v Košiciach (KEMT-TUKE) bol navrhnutý monolitický čip obsahujúci dvojicu na mieru navrhnutých zmiešavačov. Zmiešavače sú založené na konvenčne známom zapojení dvojito vyváženej Gilbertovej bunky (Obr. 3). Spotreba oboch zmiešavačov je 75 mA pri napájacom napätí -3.3 V. Navrhnuté zmiešavače majú konverzný zisk 6 dB v diferenciálnom zapojení, CMRR lepšie ako 35 dB v frekvenčnom pásme do 10 GHz a 1 dB kompresný bod v smere RF-IF je na úrovní -12 dBm. Vstupné prispôsobenie (parameter S_{11}) je v pásme do 10 GHz lepšie ako -20 dB.

Čip so zmiešavačmi je navrhnutý a vyrobený v 0.35 μm SiGe BiCMOS technológií. Celkové rozmery silikónového substrátu sú 850x1950 μm. Čip je nakontaktovaný do 5x5 mm veľkého puzdra QFN 32.

B. Výkonový delič

Hlavnou úlohu výkonového deliča je rozdeliť výkon vstupného signálu LO do viacerých signálových ciest. Výkonový delič v navrhovanom I-Q demodulátore slúži na rozdelenie signálu medzi LO vstup zmiešavača I zložky a medzi 90° oneskorovaciu linku. Realizácia výkonového deliča je možná rôznymi spôsobmi, napríklad pomocou vlnovodov, mikropásikových vedení či pasívnych súčiastok. Vzhľadom k rozmerovým dispozíciám a jednoduchosti návrhu bolo vybraté

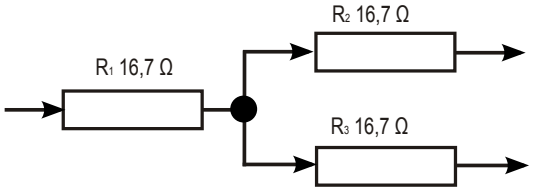


Obr. 3 Principiálna schéma zmiešavača

riešenie širokopásmového výkonového deliča pomocou rezistorov [4]. Výkonový delič pozostáva z troch odporov, ktorých hodnoty je možné zistiť pomocou vzťahu:

$$R = Z_0((N-1)/(N+1)) \tag{1}$$

kde Z_0 je charakteristická impedancia systému a N je počet ciest, do ktorých sa vstupný signál delí. Pre charakteristickú impedanciu systému rovnú 50 Ω a dvojcestný delič výkonu je výsledný odpor jednotlivých rezistorov rovný približne 16,7 Ω (Obr.4).



Obr. 4 Schéma zapojenia výkonového deliča

C. Oneskorovacia linka

Návrh oneskorovacej linky sa tiež dá realizovať viacerými spôsobmi. Jedným z možných riešení môže byť neuniformne prenosové vedenie, ktorého šírka sa dá vypočítať pomocou rozloženia do skráteného fourierového radu. Výber vhodných koeficientov sa dá dosiahnuť pomocou vhodnej optimalizácie a tým vytvoriť oneskorovacie vedenie s nízkym koeficientom odrazu v požadovanom frekvenčnom pásme [5]. Vzhľadom k tomu, že takýto výpočet je dosť zložitý a realizácia takéhoto vedenia relatívne náročná, pre návrh sme zvolili jednoduchšie riešenie pomocou tzv. mikropásikového vedenia[6]. Takéto vedenie je možné pomerne jednoducho realizovať na doske plošných spojov, jediným z problémov je určenie správnej dĺžky tohto vedenia. Dĺžku mikropásikového vedenia (Δl) je možné vyjadriť zo vzťahu (2), ktorý je daný ako pomer medzi fázovým posunom v radiánoch ($\Delta\Theta$) a fázovou konštantou (β)

$$\Delta l = \frac{\Delta\Theta}{\beta} \tag{2}$$

Fázovú konštantu vyjadríme:

$$\beta = \frac{2\pi}{\lambda_G} \tag{3}$$

kde λ_G je vlnová dĺžka substrátu. Tá je definovaná ako funkcia vlnovej dĺžky vo vzduchu (λ_{air}) v pomere s relatívnou permitivitou daného substrátu (ϵ_r), teda vzťahom:

$$\lambda_G = \frac{\lambda_{air}}{\sqrt{\epsilon_r}} = \frac{\frac{c}{f}}{\sqrt{\epsilon_r}} \tag{4}$$

kde c je rýchlosť svetla vo vákuu a f je požadovaná frekvencia, ktorej fázový posun chceme realizovať. Zo vzťahov vyplýva, že dĺžka mikropásikového vedenia sa zmenšuje s narastajúcou pracovnou frekvenciou [6].

Daná oneskorovacia linka bola realizovaná na substráte od firmy ROGERS s typovým označením RO4003C s relatívnou permitivitou ϵ_r 3,55. Pre fázový posun s hodnotou 90° a s frekvenciu LO signálu 7 GHz bola vypočítaná dĺžka mikropásikového vedenia 0,005687 m.

D. Filtre

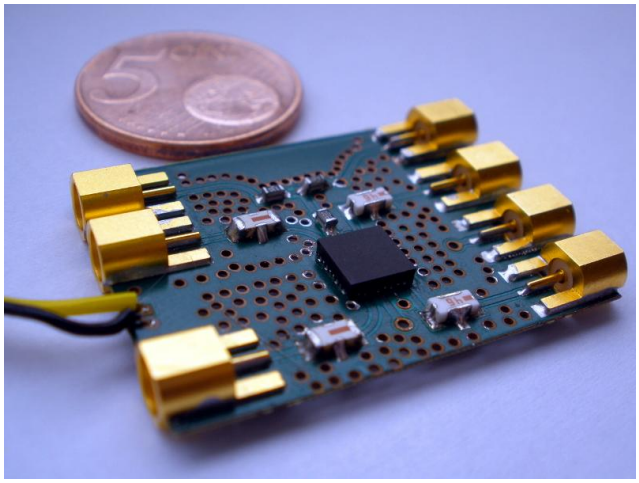
Na vstupe a výstupe I-Q demodulátora je potrebné umiestniť filtre, ktoré odstránia neželané frekvenčné zložky vstupných a výstupných signálov. Na vstupe demodulátora je umiestnená dvojica pásmových priepustov, ktorých úlohou je výber požadovaného frekvenčného pásma. Ideálnym riešením v tomto prípade by mal byť pásmový priepust s frekvenčným pásmom od 6 GHz do 8 GHz. Pre jednoduchosť sa zvolilo riešenie, ktoré zahŕňa kúpu komerčne vyrábaných filtrov. Po skúmaní trhu v tejto oblasti nebolo nájdené ideálne riešenie. Preto v prvom návrhu na overenie funkčnosti I-Q demodulátora boli ako kompromis vybrané komerčne vyrábane pásmové priepusty od firmy Mini-Circuits s označením BFCN-7900+, ktorých frekvenčný rozsah je od 7,8 GHz do 8,1 GHz.

Na výstupe I-Q demodulátora je potrebné odfiltrovať nežiadúce produkty po zmiešavaní či iné rušivé signály z vyšších frekvenčných pásiem, ktoré by negatívne ovplyvňovali digitalizovanie užitočného signálu. Bol zvolený rovnaký postup, kúpa komerčne vyrábaného dolno-priepustného filtra od firmy Mini-Circuits s označením LFCN-3400+ s frekvenčným pásmom od DC po 3,4 GHz.

Po vyššie spomínanom zistení nedostupnosti vhodného riešenia na trhu, sme zvolili riešenie založené na vlastnom návrhu potrebných filtrov. V spolupráci s Katedrou technológií v elektronike, Technickej univerzity v Košiciach sme začali návrh dolno-priepustného filtra na báze diskrétnych SMD súčiastok. Prvé vzorky navrhnutého eliptického filtra 6 rádu s hraničnou frekvenciou 2,5 GHz boli úspešne vyrobené a odmerané [7]. Tiež sa začal vývoj riešenia pásmovej priepuste s frekvenčným pásmom od 6 GHz do 8 GHz na báze LTCC keramiky [8].

IV. Realizácia

Prvá vzorka I-Q demodulátora bola vyrobená na obojstrannej doske plošných spojov na substráte RO4003C od firmy Rogers s relatívnou permitivitou 3,55 a hrúbkou 0,203 mm (Obr. 5). Tento substrát bol zvolený pre jeho vynikajúce RF vlastnosti.



Obr. 5 Fotografia prvej vyrobenej vzorky I-Q demodulátora

V. Záver

V príspevku bol uvedený prvotný návrh a realizácia I-Q demodulátora určeného na rozšírenie UWB senzorového systému založeného na pseudonáhodnej M - postupnosti. V ďalšom období bude prvá prototypová vzorka podrobená meraniam a porovnaním simulovaných a nameraných údajov bude vyhodnotená úspešnosť návrhu. Pri návrhu filtrov pre I-Q demodulátor začala tiež vzájomná spolupráca s Katedrou technológií v elektronike, Technickej univerzity v Košiciach pri návrhu filtrov nielen pre túto UWB aplikáciu. Po meraniach a ich vyhodnotení bude snaha tento rozširujúci modul k UWB senzorovému systému integrovať na LTCC keramiku.

POĎAKOVANIE

Autori ďakujú Martinovi Kmecovi a celému tímu z Ústavu Elektrického merania (FG Elektronische Messtechnik), Technickej Univerzity v Ilmenau, Nemecko za odbornú a materiálnu pomoc. Autori tiež ďakujú Kornelovi Rumanovi a tímu na Katedre technológií v elektronike, Technickej univerzity v Košiciach za návrh a výrobu dolno-priepustného filtra. Táto práca bola vytvorená realizáciou projektu Rozvoj Centra informačných a komunikačných technológií pre znalostné systémy (kód ITMS projektu: 26220120030) na základe podpory operačného programu Výskum a vývoj financovaného z Európskeho fondu regionálneho rozvoja.

LITERATÚRA

[1] Telekomunikačný úrad Slovenskej republiky, "Všeobecné povolenie číslo VPR - 02/2011," *Dostupné na internete:*, 10.3.2012, http://www.teleoff.gov.sk/data/files/22711.pdf.

[2] J. Sachs and P. Peyerl, "A New Principle for Sensor-Array-Application," *Proceedings of 16th IEEE Instrumentation and Measurement Technology Conference*, pp. 1390–1295, 1999, IMTC/99 Venice, Italy.

[3] J. Sachs, M. Kmec, R. Zetik, P. Peyerl, and P. Rauschenbach, "Ultra wideband radar assembly kit," *Geoscience and Remote Sensing Symposium*, 2005, IGARSS '05. Proceedings. 2005 IEEE International.

[4] Broadwave technologies, "POWER DIVIDERS," *Dostupné na internete:*, 10.3.2012, http://www.broadwavetechnologies.com/CataloguesPDF/application_notes_power_dividers.pdf.

[5] M. K. Amirhosseini, "Wideband differential phase shifter using microstrip nonuniform transmission lines," *Progress In Electromagnetics Research Letters„ Vol. 3*, pp. 151 – 160, 2008.

[6] RF, RFIC and Microwave Theory, Design, "Phase Shifter Design Tutorial," *www.rfic.co.uk, Dostupné na internete::*, 10.3.2012, http://www.odyseus.nildram.co.uk/RFMicrowave_Circuits_Files/Phase_%20Shifter.pdf.

[7] A. Pietriková, K. Ruman, I. Vehec, and P. Galajda, "Design and Comparison of Different Methods of Measurement of Low Pass Filter for Ultra Wide-Band Application," *Electrical Engineering and Informatics III Proceeding of the Faculty of Electrical Engineering and Informatics of the Technical University of Košice*, 2012, publikované v tomto zborníku.

[8] K. Ruman, O. Kováč, and I. Vehec, "Design of Microstrip Hairpin Band Pass Filter for UWB Application," *SCYR 2012 - 12th Scientific Conference of Young Researchers – FEI TU of Košice*, 2012, Herľany, Bude publikované.

A/D prevodník pre UWB senzorové systémy

Martin LIPTAJ, Tomáš HRAŠOK, Pavol GALAJDA

Katedra elektroniky a multimediálnych telekomunikácií, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

martin.liptaj@tuke.sk, tomas.hrasok@student.tuke.sk, pavol.galajda@tuke.sk

Abstrakt **– Článok sa venuje návrhu 5-bitového paralelného A/D prevodníka pre UWB senzorové systémy. V článku sú prezentované výsledky simulácií a meraní prevodníka navrhnutého na Katedre Elektroniky a Multimediálnych Telekomunikácií. Hlavnú časť výsledkov simulácií tvorí analýza komparátorov, ktoré predstavujú najväčší zdroj chýb v A/D prevodníkoch. Druhá časť článku prezentuje výsledky meraní a porovnáva ich s vykonanými simuláciami. V závere článku je naznačený ďalší postup integrácie navrhnutého A/D prevodníka spolu s ďalšími modulmi v jednom puzdre, ktoré predstavuje kompletný komunikačný UWB radový systém.**

Kľúčové slová **– A/D prevodník, SiGe BiCMOS, UltraWide-Band**

I. Úvod

Neustálym zdokonaľovaním výrobných technologických procesov je umožnené vytváranie čoraz menších a výkonnejších senzorových systémov pracujúcich v centimetrovom a milimetrovom frekvenčnom pásme, tzv. UWB (UltreWide-Band) pásme. Tieto systémy umožňujú široké nasadenie v oblastiach medicíny, stavebníctva, poľnohospodárstva, v automobilovom priemysle, armáde a mnohých ďalších oblastiach. Jedným z možných riešení takéhoto senzorového systému je senzorový systém pracujúci s M-postupnosťou [1], na vývoji ktorého sa podieľa aj Katedra Elektroniky a Multimediálnych Telekomunikácií, Fakulta elektrotechniky a informatiky, Technickej univerzity v Košiciach (KEMT-TUKE). Úlohou tohto vývoja je návrh softvérových ako aj hardvérových častí tohto senzorového systému a následné spracovanie a vyhodnotenie dát v reálnom čase.

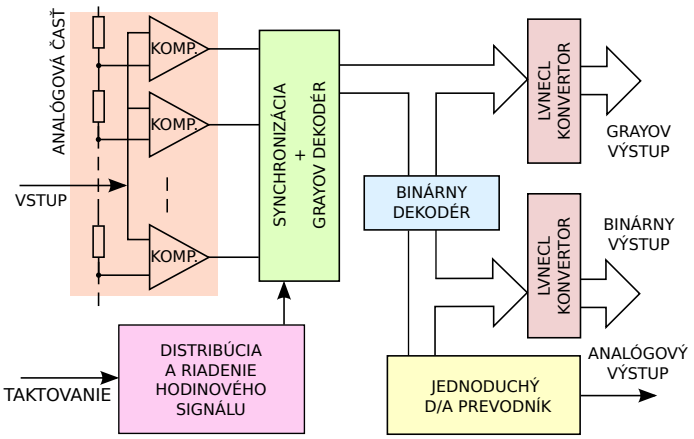
Jedným z problémov pri vývoji UWB aplikácií je korektný príjem a spracovanie prijímaného signálu. Spracovanie prijímaného signálu je najefektívnejšie realizovať číslicovými metódami spracovania signálov, preto je potrebné vstupný prijímaný analógový signál zdigitalizovať pomocou A/D (analógovo/digitálneho) prevodníka. Tento článok opisuje návrh, simulácie a merania 5-bitového paralelného A/D prevodníka navrhnutého na KEMT-TUKE.

Návrh A/D prevodníka bol uskutočnený s ohľadom na skutočnosť, že spolu s ďalšími blokmi UWB senzorového systému [2] bude integrovaný do jedného puzdra, tzv. System in Package (SiP). Takýmto spôsobom je možné eliminovať nežiaduce vplyvy puzdier a prepojení na doskách plošných spojov ako sú oneskorenie signálov, vplyv okolia, alebo tiež teplotné rozdiely a tým zvýšiť výkon celého systému.

II. Návrh A/D prevodníka

A. Štruktúra A/D prevodníka

V UWB zariadeniach pracujúcich na princípe pseudonáhodnej (PN) postupnosti dát je potrebné digitalizovať vysokofrekvenčné signály s frekvenciou rádovo stoviek MHz. Pre spracovanie takýchto signálov sú najvhodnejšie paralelné prevodníky, ktoré poskytujú veľmi rýchle digitalizovanie analógového signálu. Tieto prevodníky nazývané aj „flash“ prevodníky dokážu v jednom hodinovom takte navzorkovať vstupný signál a vygenerovať celé výstupné binárne slovo. Ich funkcia je založená na porovnávaní úrovne vstupného signálu s referenčnými napätiami na rezistorovej sieti (obr. 1). Pri zvyšovaní rozlíšenia prevodníka (počtu rozhodovacích úrovní) rastie exponenciálne počet rezistorov v rezistorovej sieti ako aj počet komparátorov a logických obvodov. Z toho vyplýva najväčšia nevýhoda paralelného prevodníka, ktorou je zložitosť návrhu vnútornej štruktúry, resp. rozmiestnenia prvkov na čipe (tzv. layout). Na rozlišovaciu schopnosť N-bitov je potrebných presne $2^N - 1$ komparátorov [3]. Aj napriek tejto nevýhode sú paralelné prevodníky často využívané hlavne v oblastiach vysokofrekvenčného pásma. Pre potreby UWB



Obr. 1 Bloková schéma navrhnutého A/D prevodníka

radaru vyvíjaného na KEMT-TUKE bol zvolený návrh 5-bitového paralelného prevodníka. Tento návrh je kompromis, ktorý zohľadňuje použitú technológiu výroby čipu, výkonovú spotrebu, náročnosť návrhu a požadované rozlíšenie prevodníka. Na základe použitej technológie bolo zvolené záporné napájacie napätie A/D prevodníka -3,3 V. Bloková schéma navrhnutého A/D prevodníka je zobrazená na obr. 1. Analógová časť prevodníka pozostáva z odporového deliča predstavujúceho zdroj referenčných napätí a 31 komparátorov. Na výstup každého komparátora je pripojený D-klopný obvod, ktorý spolu s hodinovým signálom funguje ako dočasná pamäť (latch). Tento obvod zabezpečuje synchronizáciu výstupných stavov komparátorov. Ako primárny výstupný kód prevodníka bol zvolený grayov kód. Jeho výhodou oproti binárnemu kódu je, že pri zmene o jednu referenčnú úroveň sa mení vždy iba jeden výstupný bit. Na výstup dekodéra z termometrického do grayovho kódu bol ešte pripojený dekodér z grayovho do binárneho kódu. Výstupy z dekodérov grayovho a binárneho kódu sú vybavené napäťovými konvertormi, ktoré prispôsobujú výstupné logické úrovne štandardom nízko-napäťovej negatívnej emitorovo viazanej logiky (v angličtine Low-Voltage Negative Emitter-Coupled Logic, LVNECL). Súčasťou štruktúry A/D prevodníka je aj jednoduchý D/A prevodník pozostávajúci z R-2R rezistorovej siete pripojený na výstup dekodéra z grayovho na binárny kód. Tento prevodník generuje analógový signál, pomocou ktorého je možné jednoducho zistiť funkčnosť a odhadnúť správanie sa a prípadné chyby vnútorných štruktúr A/D prevodníka.

B. Technológia výroby

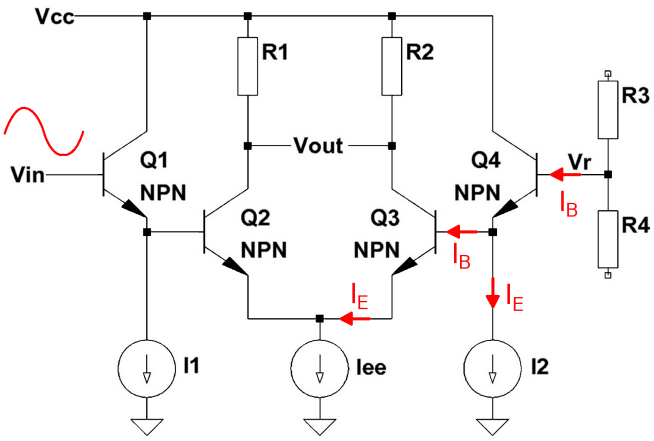
Existuje viacero spôsobov akými sa vyrábajú monolitické ASIC (application specific integrated circuit) obvody určené pre širokopásmové aplikácie. V praxi je najčastejšie uplatňovaný proces vytvárania čipu na kremíkovom (Sicilium, Si) substráte, pričom na jednom substráte je možné súčasne vyrobiť veľké množstvo rôznych čipov. Kremíkový substrát má oproti iným používaným materiálom (GaAs, InP) dobré tepelné vlastnosti a pri veľkých výrobných sériach prináša lacný spôsob výroby čipov. V procese výroby je možné kremíkový substrát dotovať pomocou prímesí kladne aj záporne, čím je možné súčasne vytvoriť polovodiče typu-n a typu-p. Tie sú potrebné pri vytváraní komplementárnych dvojíc tranzistorov PNP a NPN, ako aj pri vytváraní CMOS štruktúr. Spojením CMOS výrobného procesu s výhodami SiGe štruktúr vzniká technológia SiGe BiCMOS, ktorá zachováva nízke výrobné náklady CMOS architektúry a umožňuje realizovať aj vysokofrekvenčné štruktúry tvorené hetero bipolárnymi tranzistormi (HBT) [4]. Takého analógové obvody tvorené modernými vysokofrekvenčnými HBT môžu pracovať v oblastiach aj nad 300 GHz. Zdokonaľovanie CMOS technológie vhodnej v digitálnom spracovaní signálov je realizované zmenšovaním vrstvy dielektrika medzi riadiacou elektródou a kanálom unipolárnych tranzistorov. S týmto je spojené aj znižovanie pracovného napätia a zmenšovanie rozmerov tranzistorov, ktoré vedie k redukcii ich prúdovej spotreby a ceny [5],[3]. Výhody uvedenej SiGe BiCMOS technológie predurčujú jej hromadné využívanie v ASIC obvodoch určených pre použitie UWB radarových systémoch. Tieto ASIC obvody požadujú rýchle analógové vstupno/výstupné periférie a rozsiahle digitálne CMOS štruktúry potrebné pre spracovanie dát. Návrh uvedeného 5-bitového A/D prevodníka bol realizovaný 0.35 μm SiGe-BiCMOS technológiou poskytnutou firmou Austriamicrosystems (AMS) [6]. Táto technológia poskytuje SiGe HBT tranzistory pracujúce v oblastiach do 70GHz, pričom technológia dovoľuje realizovať systémy s priepustnosťou do 20 Gb/s.

C. Aspekty návrhu prezentovaného A/D prevodníka

Pri návrhu paralelných A/D prevodníkov sa kladie najväčší dôraz na vzorkovací obvod a komparátory. Úlohou vzorkovacieho obvodu je zachytiť aktuálnu hodnotu vstupného signálu a držať si túto hodnotu počas celej doby kvantovania signálu do kvantovacích úrovní. Navrhnutý A/D prevodník neobsahuje vzorkovací obvod ako je to pri paralelnej architektúre prevodníkov bežné. Absenciu vzorkovacieho obvodu na čipe prevodníka je možné zdôvodniť tým, že vzorkovací obvod je súčasťou UWB senzorového systému, pre ktorý je prevodník primárne určený. Na elimináciu snímania prechodových dejov vznikajúcich vo vzorkovacom obvode, je na čipe prevodníka navrhnutá oneskorovacia linka. Táto oneskorovacia linka, pripojená na taktovacie obvody, má oneskorenie 16 ns, čo predstavuje štvrtinu periódy predpokladanej taktovacej frekvencie prevodníka. Prevodník obsahuje aj priame vstupy pre taktovací signál, pomocou ktorých je možné prevodník taktovať aj bez použitia tejto oneskorovacej linky.

Vstupný signál A/D prevodníka je porovnávaný s referenčnými napätiami, ktoré sú realizované priamo na čipe pomocou odporovej siete. Táto odporová sieť je napájaná zdrojom napätia o hodnote -860 mV a tvorí ju 30 za sebou zapojených rezistorov. Na každom rezistore vzniká úbytok 29 mV. Z toho vyplýva, že každý komparátor musí byť schopný preklápať svoj výstup pri rozdiele medzi vstupným a referenčným napätím menším než 29 mV. Komparátor je realizovaný diferenciálnym zapojením bipolárnych tranzistorov, ktoré sú napájané zdrojom konštantného prúdu. Takéto zapojenie diferenčného stupňa, okrem výhodných vlastností ako vysoká citlivosť a nízka kapacita vstupných pinov so sebou prináša aj nevýhody v podobe spätného šumu (kickback noise), AC a DC odchýlke a nutnej stabilizácii pracovného bodu bipolárnych tranzistorov [3].

Keďže vstupný odpor komparátorov nie je nekonečný, je samozrejmé že do komparátora tečie malý, ale nezanedbateľný prúd z rezistorovej siete referenčných napätí. Tento prúd, ktorý vnáša odchýlky do referenčných napätí, nazývame tiež jednosmerná (DC) odchýlka. Túto odchýlku je možné korigovať správnym návrhom odporového deliča a doplnením filtračných kondenzátorov k referenčným napätiam. Princíp vzniku DC odchýlky vidíme na obrázku 2.

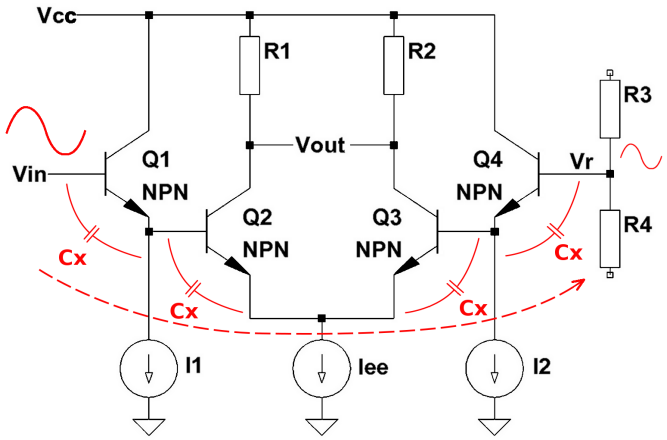


Obr. 2 DC odchýlka (bowing) komparátora

V prípade použitia paralelného A/D prevodníka v oblasti vysokých frekvencií v oblasti desiatok a stoviek megahertzov sa k DC odchýlke pripočítava ešte aj nelineárna AC odchýlka. Táto vzniká na parazitných kapacitách tvorených samotnou vnútornou štruktúrou tranzistorov a kapacitou prívodných vodičov. Na obrázku 3 sú znázornené parazitné kapacity vo vnútornej štruktúre komparátora tvoreného bipolárnymi tranzistormi.

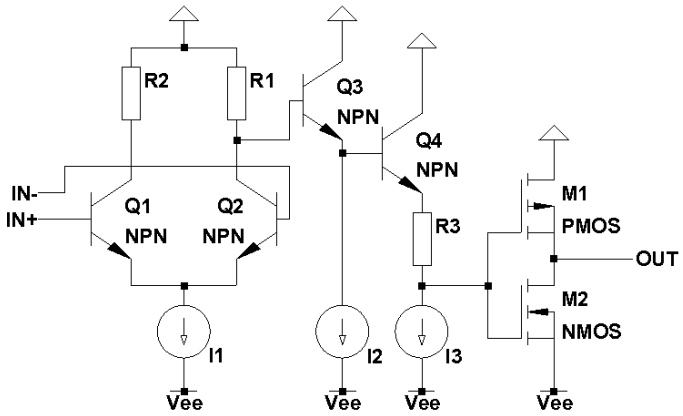
I keď bipolárne tranzistory majú pomerne nízku vstupnú kapacitu, je zrejmé, že zvyšovaním presnosti prevodníka, zväčšovaním počtu komparátorov sa priamo úmerne zvyšuje vstupná parazitná kapacita komparátora medzi vstupmi V_{in} a V_r. To vedie k prenosu vstupného signálu spätne do referenčných napätí a značným odchýlkam v referenčných napätiach. Jedným z možných riešení je vytvorenie „tvrdšieho“ odporového deliča, čo ale so sebou prináša aj vyššiu prúdovú spotrebu prevodníka. AC odchýlku je možné čiastočne korigovať aj filtračnými kondenzátormi na každom z referenčných napätí [7].

Je všeobecne známe, že vlastnosti polovodičových prvkov sa menia s teplotou. Napätie V_{BE} bipolárnych tranzistorov klesá o 2mV pri zvýšení teploty polovodiča o 1°C. Zvýšením teploty sa zvyšuje prúd bázou a kolektorom v bipolárnych tranzistoroch. Výsledkom je rozladenie prúdových zdrojov a pracovných bodov tranzistorov. Na obrázku 4 môžme vidieť blokovú schému navrhnutého komparátora. Pri simulácii zmeny teploty dochádza k nárastu v zdroji prúdu a posunu



Obr. 3 AC odchýlka (bowing) komparátora

pracovného bodu tranzistorov Q_1 až Q_4. Tento posuv sa pri teplote nad 55°C dostáva mimo rozsah rozhodovacích úrovní unipolárnych tranzistorov M_1 a M_2 a zapríčiňuje nefunkčnosť komparátora. Tento problém sa prejavil aj pri simuláciách so zmenou napájacieho napätia. Zlepšenie správania sa komparátora pri teplotných a napäťových zmenách je možné dosiahnuť stabilnejším zdrojom prúdu a zvýšením zosilnenia bipolárnej časti komparátora.



Obr. 4 Blokový diagram navrhnutého komparátora

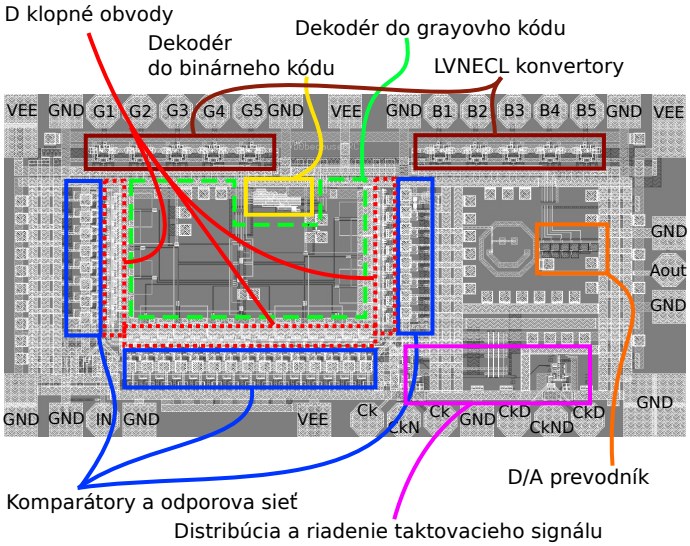
Všetky bloky navrhnutého prevodníka okrem komparátora vykazovali v simuláciách korektnú činnosť a splňovali definované požiadavky na rýchlosť a stabilitu.

D. Rozmiestnenie prvkov na čipe

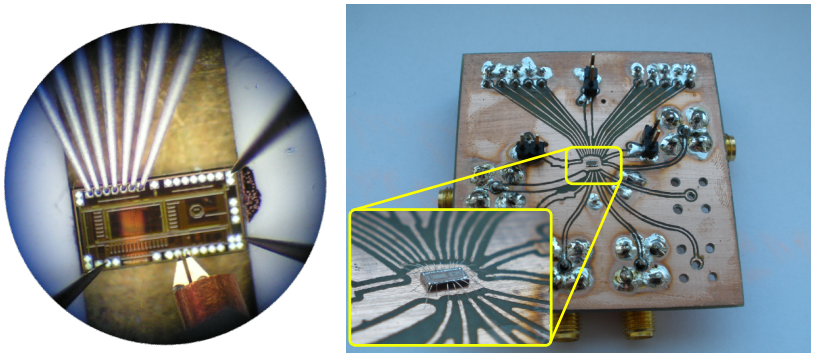
Na výrobu prevodníka bola použitá vyššie spomínaná technológia 0.35 μm SiGe BiCMOS od Austriamicrosystems, Rakúsko [6]. Veľkosť výsledného substrátu, na ktorom bol realizovaný čip A/D prevodníka je 1950x950 μm (Obr. 5).

III. Meranie prevodníka

Funkčnosť a parametre realizovaného A/D prevodníka boli overené meraniami. Prvotné merania boli realizované meraním na substráte pomocou vysokofrekvenčných meracích sond Cascade DCQ-07 a PicoProbe GS150 (obr. 6. vľavo). Kvôli priestorovým možnostiam merania na čipe bol meraný čip pripojený na napájací zdroj iba v jednom bode. Pri napájacom napätí zdroja 3,86 V bola nameraná spotreba 132 mA a bola overená základná funkčnosť prevodníka. V dôsledku nestability napájacieho napätia spôsobenej relatívne vysokým prechodovým odporom testovacích mikro-sond a nemožnosti pripojiť ďalšie napájacie sondy na eliminovanie tohto javu, nebolo možné získať relevantnejšie výsledky meraní. Pre ďalšie merania prevodníka bola preto navrhnutá testovacia doska plošných spojov (DPS), na ktorú bol čip prevodníka nakontaktovaný pomocou hliníkových vodičov (obr. 6. vpravo). Kontaktovanie ultrazvukovou metódou zaručuje stabilné a nízke prechodové odpory medzi čipom a testovacou DPS. Súčasťou testovacej dosky sú štyri



Obr. 5 Layout realizovaného prevodníka



Obr. 6 Meranie na čipe (vľavo) a osadená doska plošných spojov s nakontaktovaným A-D prevodníkom (vpravo)

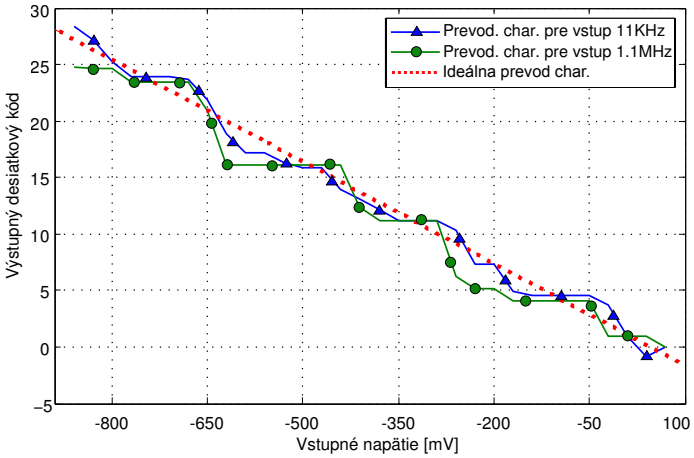
nezávislé napájacie napäťové vetvy, pomocou ktorých je možné individuálne nastaviť jednotlivé napájacie napätia pre elimináciu prípadných prechodových odporov na jednotlivých vodičoch. Tým je možné dosiahnuť vyššiu presnosť nameraných výsledkov.

Meranie prevodníka bolo vykonané pomocou histogramovej metódy a meraním s neprebudeným signálom. Na obrázku 7 sú zobrazené prevodové charakteristiky meraného A/D prevodníka pri vstupnom sínusovom signále s frekvenciami 11 kHz s taktovacou frekvenciou 10 MHz a 1,1 MHz s taktovacou frekvenciou 50 MHz. Z histogramovej metódy boli okrem prevodových charakteristík získané aj charakteristiky diferenciálnych a integrálnych nelinearít (obr. 8). Ako z týchto charakteristík vyplýva, pri zvyšovaní frekvencie dochádzalo k výpadku viacerých výstupných kódov. Príčinou boli nestabilné výstupy komparátorov spôsobené úbytkami napájacieho napätia vo vnútornej štruktúre čipu, resp. posunom pracovného bodu bipolárnych štruktúr jednotlivých komparátorov. V prípade nízkych frekvencií je tento napäťový pokles čiastočne kompenzovaný filtračnými kondenzátormi umiestnenými priamo na čipe.

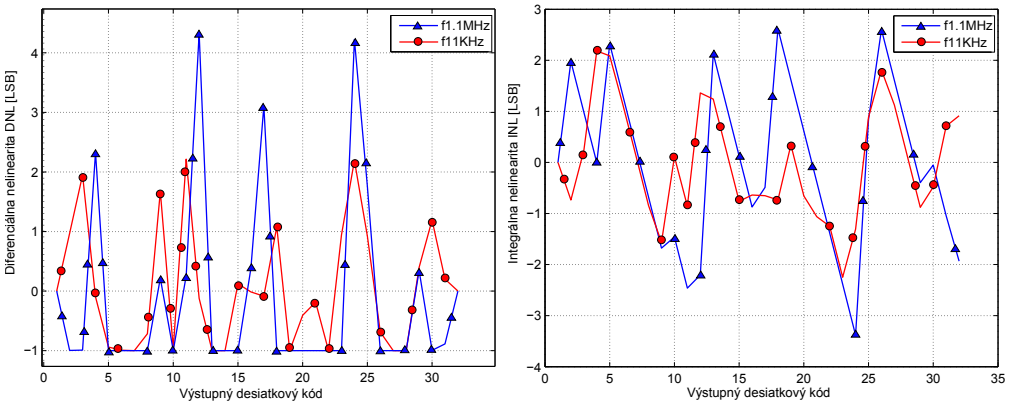
Týmito meraniami boli potvrdené výsledky simulácií, ktoré poukazujú na nestabilitu komparátorov pri zmenách teplôt a napájacích napätí.

IV. Záver

Článok sa zaoberá problematikou paralelných A/D prevodníkov so zameraním na návrh a aplikáciu 5-bitového paralelného prevodníka pomocou výrobnej technológie SiGe-BiCMOS. V skratke boli opísané vlastnosti a výhody zvolenej výrobnej technológie 0.35 μm AMS. Na základe simulácií boli popísané problémy komparátorov pri zmene teploty a napájacieho napätia. V poslednej časti článku boli zobrazené merania, ktoré potvrdzujú problematické správanie sa komparátorov. Do budúcnosti bude dôležité eliminovať vplyv napájania na funkčnosť jednotlivých blokov a tým zmenšiť nelinearity a chybovosť výstupnej charakteristiky A/D prevodníka. Potom môže byť čip



Obr. 7 Prevodové charakteristiky meraného A/D prevodníka



Obr. 8 Diferenciálne a integrálne nelinearity meraného A/D prevodníka

A/D prevodníka implementovaný spolu s ostatnými komponentami do UWB senzorového systému umiestneného v jednom puzdre prinášajúcom zvýšenie výkonu a spoľahlivosti.

Poďakovanie

Autori ďakujú Martinovi Kmecovi a celému tímu z Ústavu Eletrického merania (FG Elektronische Messtechnik) Technickej Univerzity v Ilmenau - Nemecko za odbornú, materiálnu pomoc. Autori tiež ďakujú Igorovi Vehecovi a tímu na Katedre technológií v elektronike, Technickej univerzity v Košiciach za nabondovanie čipu prevodníka. Táto práca bola vytvorená realizáciou projektu Rozvoj Centra informačných a komunikačných technológií pre znalostné systémy (kód ITMS projektu: 26220120030) na základe podpory operačného programu Výskum a vývoj financovaného z Európskeho fondu regionálneho rozvoja.

Literatúra

[1] J. Sachs and P. Peyerl, "A New Principle for Sensor-Array-Application," *Proceedings of 16th IEEE Instrumentation and Measurement Technology Conference*, pp. 1390–1295, 1999, IMTC/99 Venice, Italy.

[2] M. Kmec, M. Helbig, R. Herrmann, P. Rauschenbach, J. Sachs, and K. Schilling, "M-sequence based single chip uwb-radar sensor," *Ultra-Wideband Short-Pulse Electromagnetics 10, ANTEM/AMEREM 2010 Conference*, 2010, Ottawa, ON, Canada, july 2010.

[3] B. Razavi, "Principles of data conversion system design," 2010, ISBN 0-7803-1093-4.

[4] J. D. Cressler, "Sige technology: New research directions and emerging application opportunities," *School of Electrical and Computer Engineering , Georgia Tech, Atlanta*, Máj 2009, GA 30332-0250 USA , IEEE Southeastern Michigan Chapter IV, Ann Arbor, MI.

[5] M. Liptaj, M. Kmec, P. Galajda, and J. Sachs, "Recent sige frequency tripler development for a new uwb system architectures," *Research in Telecommunication Technologies 2010 : the 12th International Conference*, pp. 60–64, September 2010, Velké Losiny, Czech Republic. Ostrava, VŠB-TU, ISBN 978-80-248-2261-7.

[6] Austriamicrosystems, "Global leader in the design and manufacture of high performance analog ics (integrated circuits)," *Dostuné na internete: 10.03.2012:*, http://www.austriamicrosystems.com.

[7] H. Yu, "A 1v 2.5gs/s 8-bit self-calibrated flash adc in 90nm gp cmos," *University of California, Los Angeles*, 2008, ISBN 0549723951, 9780549723950.

Design texture basis of Karhunen-Loeve transform of human face

Ján MIHALÍK

Department of Electronics and Multimedia Communications, Faculty of Elektrical Engineering and Informatics, Technical University of Košice, Slovak Republic

jan.mihalik@tuke.sk

***Abstract* — This article deals about design texture basis of Karhunen-Loeve transform (KLT) of human face. We propose a way of its designing based on preprocessing training set of human faces and following PCA (Principal component analysis). This method allows it to design for human face, independent from their geometry and light conditions. Finally, we point to the possibility of its application in algorithms of digital processing of visual object - human head – for the standard MPEG-4 and MPEG-7.**

***Keywords* — texture basis, human face, preprocessing, 3D model, relevant texture analysis, MPEG-4, MPEG-7**

I. INTRODUCTION

A lot of information and communication technologies use visual information in digital form, these days. Standard video codecs [1] allow getting a digital representation of visual scenes. But codec MPEG-4 [2] unlike previous standards MPEG-1 and MPEG-2 allows coding of individual objects in video instead of whole video scenes and therefore supports the development of applications with content (object) interaction. Effective representation of objects in video, manipulation with them, editing in their bit streams as well as possibility of scaling them gives brand new ways of interaction. This creates new application areas such as: video on the Internet, mobile multimedia and communication services, content (object) access to visual databases, interactive TV, games and cinemas, etc.

SNHC (Synthetic Natural Hybrid Coding) [3] is subcategory of MPEG-4, which is specialized for coding of graphic models of real or virtual 3D objects. Standardizing of this coding extends content of original MPEG-4 applications, because it allows combining real and synthetic objects in virtual environment. Such 3D object is real human head, for which SNHC implements new algorithm of coding based on its cloning using geometric models [4], animation [5] and also texturing [6].

In the other hand, MPEG-7 [7] contains the standard tools for describing multimedia content to its fast and effective searching, indexing, filtering and access. Using MPEG-7 algorithms it is possible to make description of visual content that will inform not only about its creation, but also about conditions for its use. Also, structural information about its spatial, temporal and spatial-temporal segments is included. Information about attributes like colour, texture, motion, geometry of visual objects including composition information with mutual interaction are included, too.

A lot of algorithms of these standards using digital processing of visual object of human head is based on using texture basis of KLTof its face. We propose the method of its designing on the basis of preprocessing training set of human faces and following PCA (Principal Component Analysis), as shown in Fig.1. This method allows designing it for human face, independent from their geometry and lighting conditions.

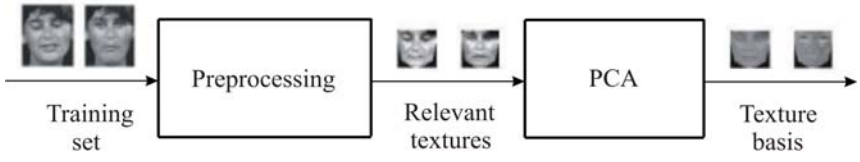
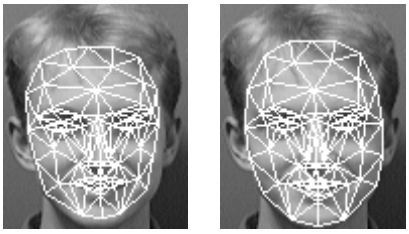


Fig. 1 Principle of designing texture basis of KLT of human face.

II. PREPROCESSING OF HUMAN FACE

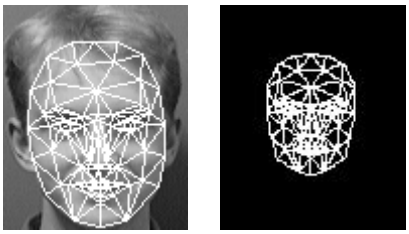
Training set of human faces contains human faces of different people, generally obtained in different light conditions. During preprocessing, 3D model [8] is adapted on human face in each image from training set using half-automated method. This method first using affine calibration [9] resizes and places this model (Fig. 2a), and then, only the position of its a few vertices is corrected manually (Fig. 2b).



a) b)

Fig. 2 a) Calibration of 3D model, b) correction of 3D model.

The next step in process of preprocessing is geometrical normalization of human face to remove texture changes made by geometrical differences among individuals. Geometrical normalization is a way of non-linear transformation which transforms source image of human face, where is allocated the adapted 3D model of human head, as seen in Fig. 3a on the standardized shape given from direct view of 3D model with a scale (Fig. 3b).



a) b)

Fig. 3 a) Original image customized with 3D model, b) Standard shape of 3D model.

After geometrical normalization of human face follows process of texturing, where barycentric coordinates [6] and scan-line algorithm [15] are used. By obtaining relevant textures from images of human faces taken in different light conditions, it is necessary to minimize changes of these light conditions before PCA application. Because of this minimizing, relevant texture is centred and energetically normalized on preprocessing output. Relevant texture values $\boldsymbol{X}$ after finishing preprocessing are

$$\boldsymbol{X} = \frac{\boldsymbol{Y} - \alpha}{\sigma} \tag{1}$$

where $\boldsymbol{Y}$ are values of not-centric relevant texture without energetic normalization. Constant α is mean value and σ standard deviation of texture. These constants are obtained from

$$\alpha = \frac{1}{\mathrm{n}} \sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{Y_i} \tag{2}$$

$$\sigma = \sqrt{\frac{1}{\mathrm{n}} \sum_{\mathrm{i}=1}^{\mathrm{n}} \left(\mathrm{Y_i} - \alpha \right)^2} \tag{3}$$

where *n* is number of elements (values) of relevant texture. All relevant textures have the standard deviation equals zero and variance equals to one, after preprocessing.

III. ANALYSIS OF RELEVANT TEXTURES

Analysis of relevant textures of human faces from training set for getting texture basis (eigen components) corresponds with Hotelling transform (HT) or Karhuen-Loeve transform [1]. It is known as Principal component analysis – PCA [10]. Result of PCA is basis of eigen components, that represents texture basis. PCA was theoretically elaborated. After that, results were applied to obtain texture basis of KLT of human face.

Generally, relevant textures are expressed in form of matrix $\boldsymbol{X}$, but for PCA needs it is better to reorganize them raw by raw (non-zero values only) to form column vector $\overline{\boldsymbol{x}}$ of size of M. Then, whole training set of relevant textures $\overline{\boldsymbol{x}}_{\mathrm{j}}$, j = 1,2,…,N can be expressed as matrix $\boldsymbol{F} = [\overline{\boldsymbol{x}}_1, \overline{\boldsymbol{x}}_2, ..., \overline{\boldsymbol{x}}_{\mathrm{N}}]$ of size M×N. It means it contains N M-dimensional column vectors of relevant textures. Then its correlation matrix is calculated as

$$\boldsymbol{C} = \frac{1}{\mathrm{N}} \boldsymbol{F}\boldsymbol{F}^{\mathrm{T}} \tag{4}$$

This matrix $\boldsymbol{C}$ of dimension M×M is symmetric and has this form

$$\boldsymbol{C} = \begin{bmatrix} c_{11}^2 & c_{12} & \cdots & c_{1\mathrm{M}} \\ c_{21} & c_{22}^2 & \cdots & c_{2\mathrm{M}} \\ \vdots & \vdots & \ddots & \vdots \\ c_{\mathrm{M}1} & c_{\mathrm{M}2} & \cdots & c_{\mathrm{MM}}^2 \end{bmatrix} \tag{5}$$

where c_{ii}^2 is own correlation moment of *i* element and c_{ij} is correlation moment between *i* and *j* elements of column vector of relevant textures, while $c_{\mathrm{ij}} = c_{\mathrm{ji}}$.

Eigenvectors $\overline{\boldsymbol{u}}_{\mathrm{i}}$, i = 1,2,…,M of correlation matrix $\boldsymbol{C}$ are calculated from

$$\boldsymbol{C}\overline{\boldsymbol{u}}_{\mathrm{i}} = \beta_{\mathrm{i}}\overline{\boldsymbol{u}}_{\mathrm{i}} \tag{6}$$

where β_{i} are eigenvalues corresponding to eigenvectors $\overline{\boldsymbol{u}}_{\mathrm{i}}$. These vectors will be columns of transform matrix $\boldsymbol{U} = [\overline{\boldsymbol{u}}_1, \overline{\boldsymbol{u}}_2, ..., \overline{\boldsymbol{u}}_{\mathrm{M}}]$, which represents texture basis of human face. As can be seen in eq. (6), calculation eigenvectors requires matrix operations, what can causes difficult calculation for big values of M. If number of N column vectors of relevant textures rapidly smaller than their dimension M, it is possible to lower this calculation by applying SVD (Singular Value Decomposition) [11]. SVD allows to express eigenvectors of matrix $\boldsymbol{C}$ as linear combination eigenvectors $\overline{\boldsymbol{g}}_{\mathrm{i}}$, i = 1,2,…,N of matrix $\boldsymbol{H} = \frac{1}{\mathrm{N}}\boldsymbol{F}^{\mathrm{T}}\boldsymbol{F}$, that can be obtained from

$$\boldsymbol{H}\overline{\boldsymbol{g}}_{\mathrm{i}} = \lambda_{\mathrm{i}}\overline{\boldsymbol{g}}_{\mathrm{i}} \tag{7}$$

where λ_{i} are eigenvalues corresponding to eigenvectors $\overline{\boldsymbol{g}}_{\mathrm{i}}$, which are columns of matrix $\boldsymbol{G} = [\overline{\boldsymbol{g}}_1, \overline{\boldsymbol{g}}_2, ..., \overline{\boldsymbol{g}}_{\mathrm{N}}]$. Then, eigenvectors $\overline{\boldsymbol{u}}_{\mathrm{i}}$ will be

$$\overline{\boldsymbol{u}}_{\mathrm{i}} = \frac{1}{\sqrt{\lambda}}\boldsymbol{F}\overline{\boldsymbol{g}}_{\mathrm{i}} \tag{8}$$

Matrix $\boldsymbol{U}$ has M eigenvectors and values, but if it is calculated using SVD, it has only N non-zero eigenvalues and vectors, the others (M-N) has value equals to zero. Calculation of eigenvectors of matrix $\boldsymbol{C}$ of dimensions M×M is reduced on calculation of eigenvectors of matrix $\boldsymbol{H}$ of dimensions N×N, assuming N < M. Decomposition of relevant texture $\overline{\boldsymbol{x}}$ from any human face, generally also outside of the training set, using texture basis designed by this way is given by forward and inverse KLT with transform matrix $\boldsymbol{U}$

$$\overline{\boldsymbol{y}} = \boldsymbol{U}^{\mathrm{T}}\overline{\boldsymbol{x}}$$
$$\overline{\boldsymbol{s}} = \boldsymbol{U}\overline{\boldsymbol{y}} \tag{9}$$

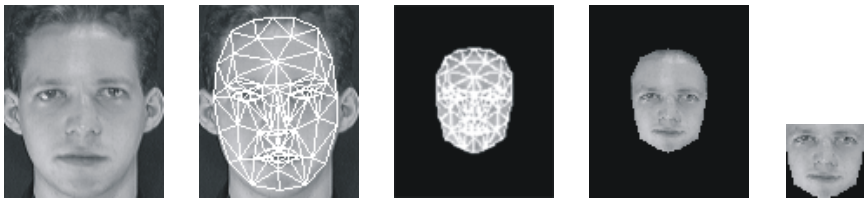
where $\overline{\boldsymbol{y}}$ represents vector of spectral coefficients and $\overline{\boldsymbol{s}}$ is column vector of approximated relevant texture. Difference between relevant texture and its approximation will be minimum, under the summed square error

$$\mathrm{SSE} = \left\|\overline{\boldsymbol{x}} - \overline{\mathbf{s}}\right\|^2 \tag{10}$$

what express square value of Euclidean norm of difference vector $\overline{\boldsymbol{x}} - \overline{\mathbf{s}}$.

IV. DESIGNED TEXTURE BASIS OF KLT OF HUMAN FACE

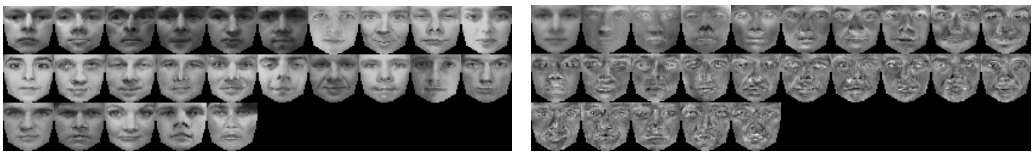
Designed texture basis of KLTof the human face is obtained from images of human faces in frontal view and no motion in the training set that captures as many differences in its geometrical shape as possible. In our experiments, we designed such a texture basis from training set of 25 different human faces in frontal view and with no motion. 3D model of human head adapted (Fig.4) using half-automated method was used first. After that, geometrical normalization of this adapted model to standard shape with direct look of 3D model human head with closed eyes and mouth with given scale (Fig. 4c) was carried out. Geometrically normalized texture on Fig. 4d corresponds to this standardized model and its relevant part was centred and energetic normalized. After full preprocessing, relevant texture on Fig. 4e has size 40x42 pels. Relevant texture obtained this way was rearranged from matrix form to column vector $\overline{\boldsymbol{x}}_{\mathrm{j}}$, that contains M = 1441 items. Finally, using PCA training set of relevant textures was obtained texture basis of KLTof human face.



a) b) c) d) e)

Fig.4 a) Original image of human face, b) original image with adapted model, c) standardized shape of model with texture, d) geometrically normalized texture, e) relevant texture.

Fig.5 shows 25 relevant textures obtained from images in training set and also complete texture basis of 25 eigenfaces.



a) b)

Fig. 5 a) Set of 25 relevant textures b) texture basis of 25 eigenfaces.

Sufficiently accurate approximation of the relevant texture of human face can be also obtained by a limited number of eigenfaces from designed complete texture basis. This reduction can reduce the computational complexity of this approximation. If the relevant texture is obtained from training set, SSE its approximation using complete texture basis of 25 eigenfaces from Fig. 5 is equal zero. On the other side, using the complete texture basis SSE value is not zero for textures obtained from images outside the training set.

The more eigenfaces is used, the less is SSE of approximated relevant texture, but it is more difficult to calculate it. As suitable compromises are reduced texture bases e.g. 10 eigenfaces from the complete texture one. For illustration, Fig. 6 shows accuracy of approximation the relevant texture 3 from "Miss America", that was obtained at first from complete and then from reduced texture basis of 10 eigenfaces. For the complete texture basis with 25 eigenfaces its

accuracy is exactly perfect with the value of SSE=0. For this reduced texture basis is accurate enough with SSE=74 and very good subjective quality,as seen in Fig.6.

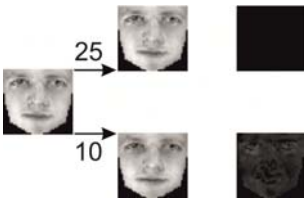


Fig. 6 (from left to right) Relevant texture, approximated relevant texture, difference pictures.

V. CONCLUSION

Video codec MPEG-4 supports coding on analysis and synthesis of human head base. Using this coding consists of its detection and following adaptation of 3D model of human head that is part of coder and also decoder. After human head detection in the first frame, technique of shaping follows. This adapts geometric shape 3D model to human head in input video sequence. Shaping of 3D model geometry in the first frame can be completly made by using the algorithm of estimating its shaping parameters [12] that needs texture basis of human face. Simple (classical) algorithms of shaping 3D model are based on extraction feature points on human face in the first frame [9], as mouth corners, nose tip and root, the centers of the eyes etc. From these points, parameters adapting geometric shape of 3D model are obtained. For this extraction can be used similar texture basis, but generated only for small parts of human face containing the feature points. Way of their design is analogical as for whole human face. They are obtained from training set of these individual parts. Motion of shaped 3D model is specified by its global and animation parameters, which can be estimated using the algorithms based on texture basis with motion [13].

Visual descriptors [14] for human face recognition in standard MPEG-7 are using KLT algorithm with designed texture basis of eigenfaces. This makes it universal and very efficient not only for recognition human faces with different texture look, but also for faces with different geometric shape and is less sensitive for light conditions changes. Sufficient efficiency of their recognition is ensured by texture basis with small number of eigenfaces, or by feature vector of spectral coefficients KLT with small dimension. Its universality is also in recognition of human faces with different geometric shape and scale, what is given by its geometrical normalization. This normalization approaches geometrical similarity of many human faces, what makes their recognition mostly by texture differences.

REFERENCES

[1] J. Mihalík, *Kódovanie obrazu vo videokomunikáciach*, Mercury-Smékal, Košice, 2001. ISBN 80-89061-47-8.

[2] J. Mihalík, "Štandardný videokodek MPEG-4, " *Slaboproudý obzor*, 2003, roč. 60, č.2, s. 7-11.

[3] The special issue of *the IEEE Trans.on Circuits and Systems for Video Technology* on MPEG-4 SNHC, July 2004.

[4] M. Strintzis – N. Sarris, (Ed.): *3D modeling and animation: Synthesis and Analysis Techniques for the Human Body*, IRM Press, Hershey, PA, July 2004.

[5] I. Pandzi – R. Forchheimer, *MPEG-4 Facial Animation: The Standard, Implementation and Aplications*, John Wiley and Sons, 2002.

[6] J. Mihalík – V. Michalčin, "Texturing of Surface of 3D Human Head Model," *Radioengineering*, Vol.13, No.4, 2004, p.44-47.

[7] T. Sikora, "The MPEG-7 Visual Standard for Content Description – An Overview," *IEEE Trans. on Circuits and Systems for Video Technology*, Vol.11, No.6, 2001, p.696-702.

[8] J. Ahlberg, *Candide-3: An Updated Parameterised Face*, Rep. No. LiTH-ISY-R-2326, January 2001.

[9] J. Mihalík – V. Michalčin, "3D Motion Estimation and Texturing of Human Head Model," *Radioengineering*, Vol.13, No.1, 2004, p.26- 31.

[10] T.F. Cootes – C.J. Taylor, *Statistical Models of Appearance for Computer Vision*, Technical report, University of Manchester, 2004.

[11] F. Dornaika - J. Ahlberg, *Fast and Reliable Active Appearance Model Search for 3D Face Tracking*, Proceedings of Model-based Imaging, Rendering, Image Analysis, and Graphical Special Effects (Mirage), 2003, p. 113–122.

[12] J. Mihalík – M. Kasár, *Shaping of Geometry of 3D Human Head Model*, Proc. 17th International Conference "Radioelektronika 2007", Brno, Czech Republic, 2007, pp. 483-486.

[13] J. Mihalík – M. Kasár, "Human Face and Facial Feature Tracking by Using Geometric and Texture Models," *Journal of Electrical Engineering,* Vol.59, No.5, 2008, p.266-271.

[14] P. Salambier, "Overview of the MPEG-7 Standard and of Future Challenges for Visual Information Analysis," *EURASIP Journal on Applied Signal Processing,* No.4, 2002, p.1-11.

[15] J.D. Foley – A. Van Dam – S.K. Feiner – J.F. Hughes, *Computer Graphics, Principles an Practicles,* Addison – Wesley, 2. Edition, 1990.

Dostupnosť hybridnej FSO/RF linky

Pavol MIŠENČÍK, Ján TURÁN, Ľuboš OVSENÍK

Katedra Elektroniky a multimediálnych telekomunikácií, Fakulta Elektrotechniky a Informatiky, Technická Univerzita v Košiciach, Slovenská Republika

pavol.misencik@tuke.sk, jan.turan@tuke.sk, lubos.ovsenik@tuke.sk

***Abstrakt* — Free Space Optics (FSO) technológia s veľkou šírkou pásma sa stala bezdrôtovou alternatívou k optickej kabeláži. Hlavnými výhodami FSO oproti vláknu je jeho rýchle nasadenie, či výrazne úspory nákladov pri inštalácii. Za nevýhodu FSO sa považuje, že prijatý výkon je variabilný a ťažko predvídateľný. Je to v dôsledku pôsobenia počasia na prenášaný optický lúč. Táto práca sa zaoberá dostupnosťou a vlastnosťami hybridnej FSO/RF linky, ktorá poskytuje vysokú dostupnosť (99,999 %), čo v praxi znamená takmer nepretržitú komunikáciu.**

***Kľúčové slová* — FSO systém, hybridný FSO/RF systém, vlastnosti hybridného FSO/RF systému, MIMO prenosový systém, FlightStrata 155E, Hybrid FSO/RF systém**

I. Úvod

FSO širokopásmová komunikačná technológia je menej známou bezdrôtovou technológiou než RF technológia. Point-to-point FSO prenosový systém môže dosiahnuť dátovú rýchlosť porovnateľnú s dátovou rýchlosťou káblových optických komunikačných systémov bez toho, aby došlo k zvýšeniu ceny a času na inštaláciu v porovnaní s vláknovou optikou. Avšak široké využitie FSO komunikačných systémov bolo sťažené vzhľadom k problémom spoľahlivosti [4]. Spoľahlivosť FSO komunikačnej linky výrazne klesá v prítomnosti dymu a hmly. V dôsledku toho sa nedosiahne požadovaná dostupnosť 99,999 %. Riešením pre dosiahnutie tejto dostupnosti je využitie RF linky ako záložnej linky k FSO linke. FSO linka trpí v dôsledku atmosférickej turbulencie zapríčiňujúcej kolísanie indexu lomu v atmosfére, pôsobením hmly a pod. [1]. To však nemá vplyv na RF linku. FSO linka poskytuje obrovské bezlicenčné pásma, kým spektrum RF linky je všeobecne licencované a obmedzené. RF linka trpí dôsledkom pôsobenia dažďových zrážok, kým útlm FSO linky je podstatne nižší pri pôsobení dažďových zrážok. Hybridný kanál môže využívať výhody oboch FSO a RF liniek. FSO linka zálohovaná RF linkou, využíva silné vlastnosti každého systému na prekonanie ich slabých stránok podrobne uvedených v [1], [5]. Takto môže hybridný FSO/RF systém dosiahnuť vyššiu dostupnosť linky, ako je dosiahnutá samostatnou FSO linkou alebo RF linkou.

II. Aplikácia a hlavné charakteristiky FSO/RF linky

Hybridnú FSO/RF prenosovú linku uvažujeme z dôvodu zálohovania primárnej FSO prenosovej linky RF linkou, ktorá môže zlyhať v dôsledku nepriaznivých vplyvov, no hlavne hmly. Tento hybridný systém pozostávajúci z dvoch liniek pracuje paralelne a RF linka nahrádza FSO primárnu linku pri jej zlyhaní. FSO linku v hybridnom FSO/RF tvorí v konkrétnom uvažovanom prípade LightPointe systém FlightStrata 155E, ktorý máme v našom laboratóriu LOS KEMT FEI TUKE k dipozícii. RF linka je vytvorená zo 60 GHz RF linky. Keďže FSO systém pracuje s prenosovými rýchlosťami do 1 Gb/sec., preto využijeme RF linku s mm vlnami, aby sme zachovali čo najviac výhod uvažovanej FSO linky.

A. FSO charakteristika

Zo všetkých nepriaznivých vplyvov na FSO linku ma najväčší vplyv hmla, ktorá môže zapríčiniť úplne zlyhanie FSO linky. V metrológii sa intenzita hmly vyjadruje viditeľnosťou. Uvediem niektoré príklady útlmu a viditeľnosti pre FSO komunikačný systém s vlnovou dĺžkou 850 nm. Ľahká hmla, ktorá predstavuje viditeľnosť okolo 10 km spôsobuje útlm FSO linky 1 až 4 dB/km. Ťažká hmla, ktorá predstavuje viditeľnosť menšiu ako 0,5 km spôsobuje veľmi veľký útlm FSO linky, ktorý je väčší ako 40 dB/km a môže dosahovať hodnoty útlmu až do 300 dB/km. Útlm spôsobený hmlou závisí aj na vlnovej dĺžke prenášaného signálu. Je menší na dlhších vlnových dĺžkach, ale rádovo je podobný.

Útlm spôsobený silným dažďom je nezanedbateľný a závisí od intenzity zrážok. Veľmi silné zrážky, čo je 90 až 100 mm/h, vyskytujúce sa v miernom klimatickom podnebí spôsobujú útlm cca 20 dB/km.

B. RF charakteristika

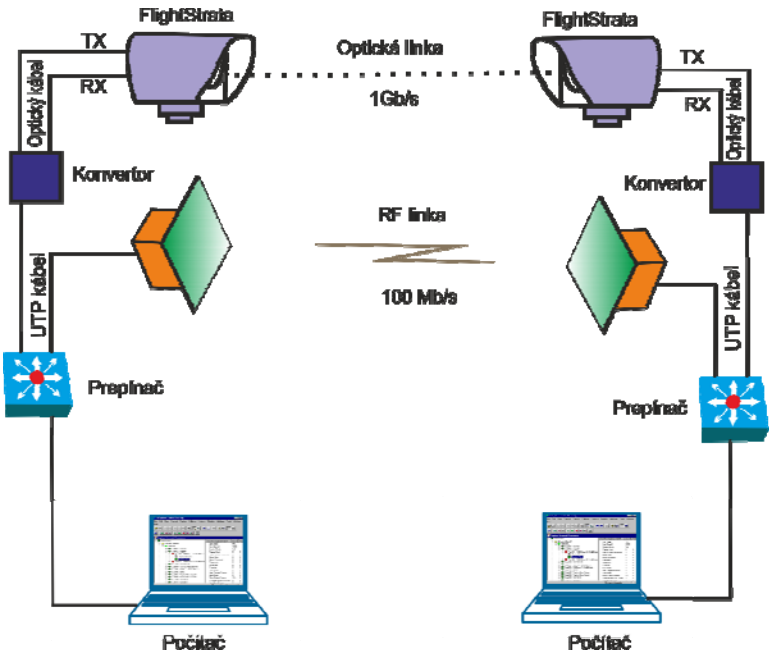
Hmla spôsobuje do istej miery straty v RF prenosovom kanáli, ale tieto straty na rozdiel od strát v FSO prenosovom kanáli sú zanedbateľné. Straty spôsobené hmlou závisia od obsahu kvapôčok vody vo vzduchu a teploty vzduchu. V porovnaní s FSO, pre mm vlny je hlavnou príčinou výpadku spojenia silný dážď. Útlm RF linky môže byť 25-35 dB/km pri zrážkach 90 mm/h a je iba slabo závislý od frekvencie.

Z hľadiska dostupnosti najvhodnejšie záložné linky by boli linky pracujúce s frekvenciu pod 10 GHz (napr. [3]), kde vplyv hmly a dažďa na prenášaný signál je malý. V skutočnosti však toto riešenie je veľmi ťažko dosiahnuteľné z rôznych dôvodov. Prenosová rýchlosť záložnej linky by musela byť aspoň 100 Mbit/s a neskôr možno 300 – 1000 Mbit/s, ktorá je nad limit aktuálnych technických možností. Jednou z veľkou výhod FSO je bezlicenčné pásmo, zatiaľ čo táto výhoda pre navrhované zálohovacie RF linky s frekvenciami pod 10 GHz je ťažko dosiahnuteľná. (aj keď to neplatí pre 2,4 a 5 GHz ISM/UNI pásma, ale preťaženie v týchto pásmach je extrémne veľké.)

Zálohovanie mm vlnami môže byť navrhované a riešené v rôznych frekvenčných pásmach. Náš model uvažuje RF linku, ktorá je tvorená 60 GHz anténou kvôli je špecifickým vlastnostiam, ktoré sú presne určené a navrhnuté v nasledujúcej kapitole. 60 GHz linka je v podstate zdarma (bez licencie), taktiež je bez vnútro systémovej interferencie a šírka pásma je dostatočne vysoká aj pre Gb/s prenosy. Na druhej strane toto frekvenčné pásmo možno použiť iba na krátke spojenie, typicky do 1 km.

C. Hybridný FSO/RF system

Hybridný FSO/RF systém zobrazený na *Obr.1* tvorí FSO systém LightPointe E 155. Technické parametre tohto systému sú zobrazená v *Tab. 1*. Pomocou týchto parametrov a údajov o viditeľnosti získaných z internetovej stránky Meteo (www.meteo.sk) určíme dostupnosť systému LightPointe E 155. Hodnoty viditeľnosti boli pomocou skriptu sťahované a ukladané na náš server v polhodinových intervaloch.



Obr. 1 Hybridný FSO/RF systém.

Záložná RF linka je tvorená 60 GHz linkou a je v prevádzke iba vtedy, ak primárna FSO linka zlyhá. 60 GHz linku sme navrhli kvôli jej výhodným vlastnostiam bližšie popísanými v [2]. Niektoré z nich sú:

- Nelicencované prevádzkovanie – nie je nutné stráviť veľa času a financií na získanie licencie od FCC.
- Vysoko bezpečná a neinterferujúca prevádzka – vyplývajúca z krátkej prenosovej vzdialenosti spôsobenej absorpciou O_2 a úzkej šírky vyžarovaného lúča.

- Možnosť viacnásobného využívania jednej frekvencie – je výhodou pri komunikácii viacerých zákazníkov v rovnakej geografickej oblasti.
- Možná optická rýchlosť prenosu dát – je k dispozícii nepretržité 7 GHz frekvenčné pásmo v porovnaní s menej ako 0,3 GHz u iných nelicencovaných pásmach.

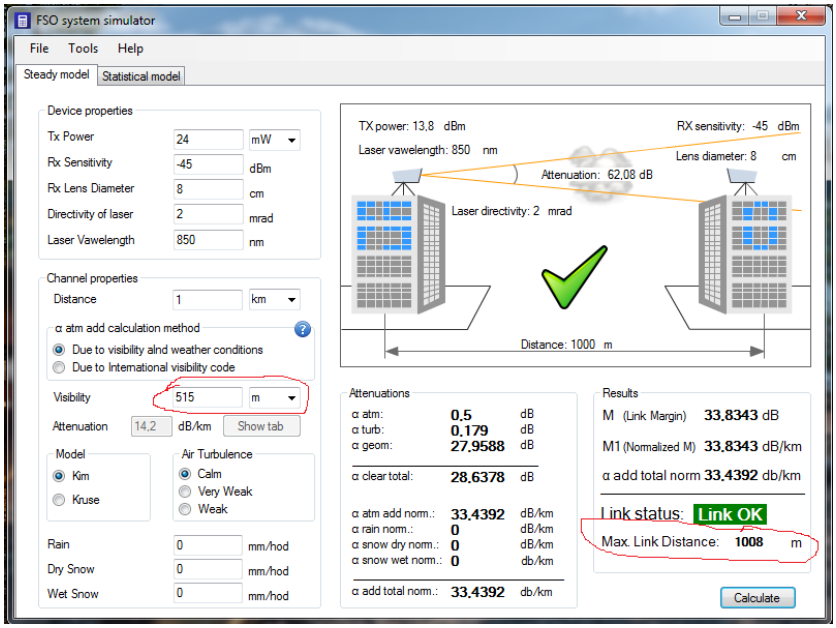
III. Dostupnosť Hybridnej FSO/RF Linky

Tabuľka 1
Technická špecifikácia FSO systému LightPointe FlightStrata 155E.

Flightstrata 155E and 155 EW	Parametre
Rozmery	11.8 x 11.8 x 25 inches
Váha	29.7 lbs / 13.5 kg
Linkhead input voltage	12 – 16 VDC
Napájacie napätie	115 / 230 VAC (50/60 Hz) or ± 48 VDC
Spotreba systému	Max. 20 W
Pracovný teplotný rozsah	-13° to +140°F (-25° to +60° C)
Šírka pásma	1.5 to 155 Mbps
Optický vysielač	VCSEL
Výstupná vlnová dĺžka	850 nm
Divergencia lúča	FSA 155E 2.0 mr FSA 155EW 2.8 mr
Vysielaný laserový výkon	FSA 155E, 155EW 24 mW
Citlivosť prijímača	FSA 155E, 155EW -45 dBm
Typ prijímača	Si APD

A. Dostupnosť FSO linky využívajúcej LightPointe FlightStrata 155E systém

Pomocou FSO System Simulator (FSO SystSim), ktorý bol navrhutý a implementovaný v LOS KEMT FEI TUKE, zistíme kedy a pri akej viditeľnosti FSO linka, ktorú tvorí LightPointe FlightStrata 155E systém, je schopná a kedy nie je schopná prenosu. Vstupné údaje pre tento systém, ktoré sú potrebné, sú v *Tab. 1*. Hlavným parametrom hybridnej FSO/RF linky je prenosová dĺžka linky. Naša hybridná linka bude operovať na vzdialenosť 1 km a pre túto vzdialenosť určíme dostupnosť hybridnej linky v Košiciach. Pomocou FSO SystSim určíme hodnotu viditeľnosti, pri ktorej FSO linka nebude schopná prenosu dát, teda vypadne. Na *Obr. 2* je výstup z programovacieho balíkaFSO SystSim. Hodnota viditeľnosti 515 m je hraničnou hodnotou, pri ktorej FSO linka prestáva pracovať. Viditeľnosť pod hodnotu 515 m zapríčiní výpadok FSO linky.

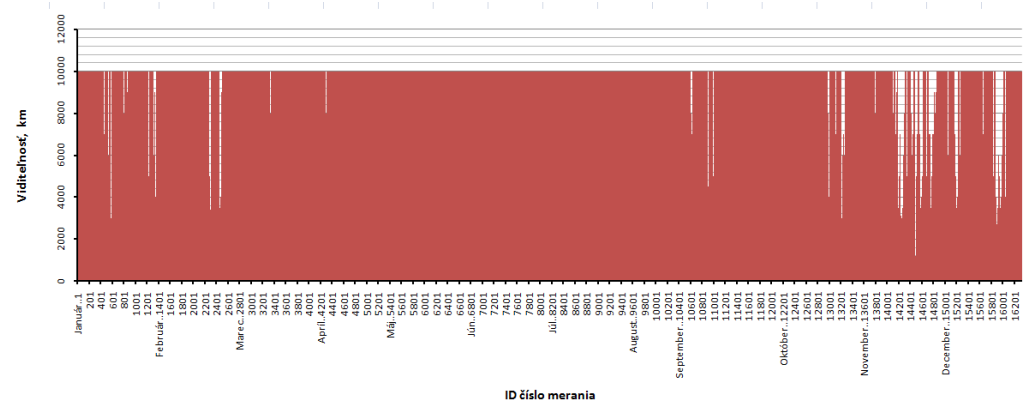


Obr. 2 FSO SystSim, Link OK.

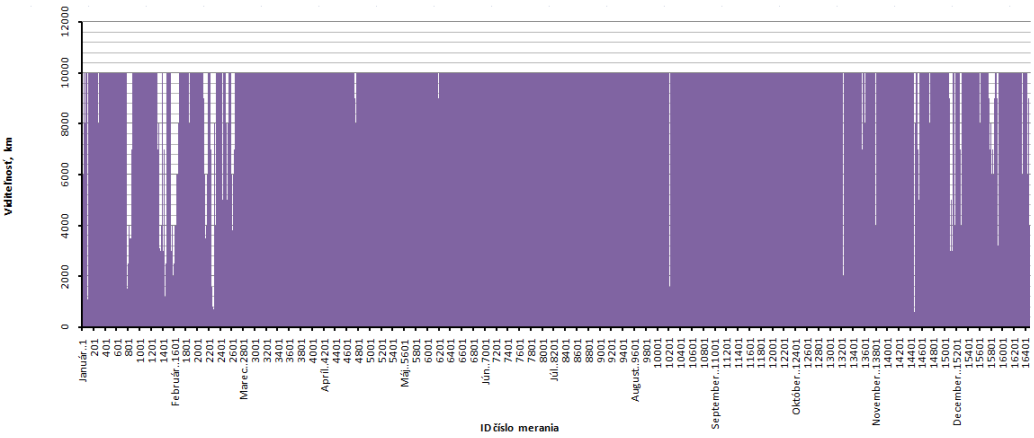
Merané hodnoty viditeľnosti za rok 2009 a 2010 sú zobrazené na *Obr.3* a *Obr.4*. Na týchto grafoch sú znázornené závislosti viditeľnosti merané v kilometroch od ID poradového čísla merania. Hodnoty sú merané v polhodinových intervaloch od 1. Januára 2009 až do 31.

Decembra 2010. Každej odmeranej hodnote bolo pridelené ID číslo kvôli lepšej manipuláciu s nameranými hodnotami. V grafoch sú vykreslené všetky namerané hodnoty, ale na x-ovej osi nie sú zobrazené všetky ID hodnoty kvôli lepšej zrozumiteľnosti. Taktiež, pri prvej hodnote ID v novom mesiaci je označený daný mesiac. Z grafu vyplýva, že merané hodnoty viditeľnosti sú od 100 m až do 50 km.

Keďže údaje sú merané v polhodinových intervaloch, tak počet meraní by sa mal rovnať dvojnásobnému počtu hodín v roku. V skutočnosti však presný počet meraní počas jedného roka sa nerovná číslu 17520, ale je o niečo menší. Je to spôsobené výpadkami servera, preto je tento počet trošku nižší. Pre výpočet dostupnosti pre oba FSO systémy budem používať presný počet meraní, ktorý pre rok 2009 je 16336 a pre rok 2010 je 16491.



Obr. 3 Viditeľnosť meraná v polhodinových intervaloch počas roka 2009.

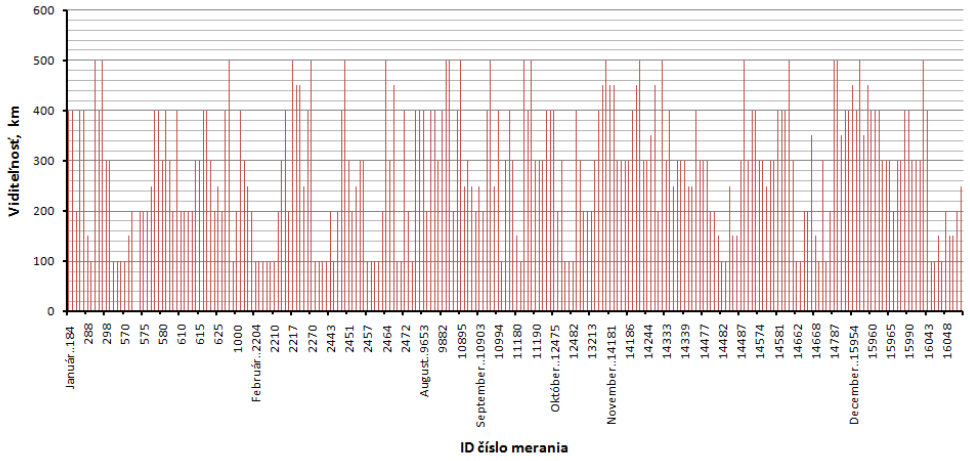


Obr. 4 Viditeľnosť meraná v polhodinových intervaloch počas roka 2010.

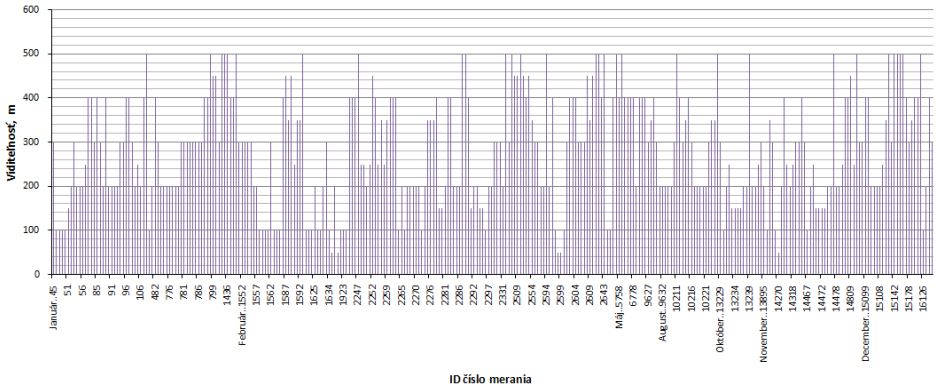
Pre lepšie určenie hodnôt viditeľnosti, menších ako 515 m nám slúži *Obr. 5 a Obr.6*. Ako na nich môžeme vidieť, najviac hodnôt, ktoré sú menšie ako 515 m, je v mesiacoch január, február, október, november a december (porovnaním ID v *Obr. 3* a *Obr. 4*). To sa dalo aj čakať vďaka tomu, že v týchto mesiacoch je viditeľnosť oveľa nižšia kvôli hmle a zrážkam.

Na určenie dostupnosti využijeme štatistické vzťahy. Celkový počet hodnôt, ktoré sú menšie ako 515 m je 240 v roku 2009. Keďže hodnoty viditeľnosti sú merané v polhodinových intervaloch, celkový počet meraní bude 16336 za rok. Dostupnosť systému FSO LightPointe FlightStrata 155E je určená ako:

$$D = \left(1 - \frac{\textit{čas výpadkov}}{\textit{celkový čas merania}}\right) * 100\% = \left(1 - \frac{240}{16336}\right) * 100\% = 98{,}5308\% \qquad (1)$$



Obr. 5 Hodnoty viditeľnosti spôsobujúce výpadok linky FSO LightPointe FlightStrata 155E systému za rok 2009.



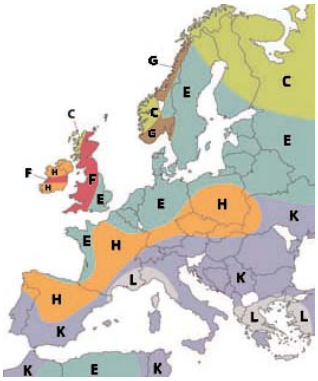
Obr. 6 Hodnoty viditeľnosti spôsobujúce výpadok linky FSO LightPointe FlightStrata 155E systému za rok 2010.

Tým istým spôsobom ako sme určili dostupnosť za rok 2009 určíme dostupnosť FSO systému aj za rok 2010. Rozdiel bude iba v tom, že celkový počet meraní bol 16491. Teda dostupnosť FSO systému LightPointe FlightStrata 155E je určená ako:

$$D = \left(1 - \frac{\textit{čas výpadkov}}{\textit{celkový čas merania}}\right) * 100\% = \left(1 - \frac{304}{16491}\right) * 100\% = 98{,}1566\% \qquad (2)$$

B. Dostupnosť RF linky

Ako RF systém využijeme AireBeam Z60 systém reprezentujúci zariadenie, ktoré poskytuje mikrovlnu technológiu na vonkajšie prepojenie dvoch bodov pripojených k sieti ethernet. Zariadenie zaručuje najvyššiu dostupnosť na kratšie a stredné vzdialenosti v typických európskych klimatických podmienkach s dosiahnutou 99,9 % pravdepodobnosťou prenosu dát na vzdialenosť takmer jedného kilometra. Z Tabuľky *Tab. 2* a *Obr. 7* [6] podrobnejšie určíme dostupnosť RF linky pre jednotlivé oblasti v Európe. Systém pracuje s frekvenciami od 59 do 63 GHz a teda podporuje aj našu bez licenčnú prevádzku 60 GHz.



Obr. 7 Klimatické oblasti pre Európu

Tabuľka 2
Spoľahlivosť prenosu a vysielací dosah v oblastiach

Oblasť / Pravdepodobnosť	E	F	H	K	M
99,999%	550	525	510	480	440
99,997%	650	600	600	550	490
99,99%	770	730	700	650	570
99,97%	870	830	800	760	660
99,90%	950	920	900	870	770
99,70%	1030	980	990	990	880
99%	1080	1050	1040	1050	990

*všetky prenosové vzdialenosti sú v metroch

C. Dostupnosť FSO/RF linky

Celkovú dostupnosť hybridnej linky určíme pomocou štatistickej metódy. Dostupnosť RF linky určíme z tabuľky Tab. 2 a to 99,7 %. V nasledujúcej *Tab. 3* sú vypočítané jednotlivé dostupnosti a taktiež celková dostupnosť hybridného FSO/RF systému počas doby dvoch rokov.

Dostupnosť vypočítaná za rok 2009 je:

$$D_{FSO/RF} = \left[1-\left(1-D_{FSO}\right)\times\left(1-D_{RF}\right)\right]\times 100\% = \left[1-\left(1-D_{FSO}\right)\times\left(1-D_{RF}\right)\right]\times 100\% = \left[1-\left(1-0{,}985308\right)\times\left(1-0{,}997\right)\right]\times 100\% = \underline{99{,}9956\%} \quad (3)$$

Tabuľka 3
Dostupnosť hybridného FSO/RF systému

Obdobie (rok)	FSO linka	RF linka	FSO/RF
2009	98,5308%	99,7%	99,9956%
2010	98,1566%	99,7%	99,9945%
Spolu(2009+2010)	98,3428%	99,7%	**99,9950%**

IV. ZÁVER

V článku sme sa zaoberali dostupnosťou hybridnej FSO/RF linky v Košiciach. Táto linky využíva systém FSO LightPointe FlightStrata 155E, ktorý máme v našom laboratóriu (LOS KEMT FEI TUKE). Z *Tab. 3* vyplýva, že dostupnosť hybridnej linky počas dvoch rokoch je 99,9950 %, ktorá je postačujúca. Nefunkčnosť tejto hybridnej linky počas dvoch rokov je iba 0,876 hodín čo je 52,56 sekúnd. Do jedného roka je to v priemere 26,28 sekúnd.

POĎAKOVANIE

Podporujeme výskumné aktivity na Slovensku/ Projekt je spolufinancovaný zo zdrojov EÚ. Tento článok bol vypracovaný v rámci projektu "Centrum excelentnosti integrovaného výskumu a využitia progresívnych materiálov a technológií v oblasti automobilovej elektroniky", ITMS 26220120055 (50%). Táto práca bola taktiež podporovaná a realizovaná z projektu Rozvoj Centra informačných a komunikačných technológií pre znalostné systémy (ITMS kód: 26220120030) na základe podpory operačného programu Výskum a vývoj financovaného z Európskeho fondu regionálneho rozvoja (50%).

LITERATÚRA

[1] M. Alnaboulsi, H. Sizun, F. deFomel, „Fog Attenuation for Optical and Infrared Waves“, *Journal of Optical Eng*., vol. 43, 2004, pp. 319-329.

[2] CH. Koh, „The Benefits of 60 GHz Unlicensed Wireless Communication“. Dostupné na internete: www.ydi.com

[3] R. Luna, D. K. Borah, H. Tapse, „Behavior of hybrid optical/RF channels…“, *Proc. IEEE GLOBECOM 2009*, paper# ONS-05/4, Honolulu.

[4] F. Nadeem, V. Kvicera, A. M. Saleem, E. Leitgeb, S.S. Muhammad, G. Kandus, „Weather effects on Hybrid FSO/RF Communication link“, *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 9, December 2009, pp. 1687-1697.

[5] R. Srinivasan, D. Sridharan, „The climate effects on line of sight (LOS) in FSO communication“, *IEEE Internationl Conference on Computational Intelligence and Computing Research*, 2010.

[6] AireBeam Z60 datasheet. Dostupné na internete: http://www.lightpointe.com/images/AireBeam_Z60.pdf

Hybridný FSO/MIMO RF prenosový systém

Pavol MIŠENČÍK, Ján TURÁN, Ľuboš OVSENÍK

Katedra Elektroniky a multimediálnych telekomunikácií, Fakulta Elektrotechniky a Informatiky, Technická Univerzita v Košiciach, Slovenská Republika

pavol.misencik@tuke.sk, jan.turan@tuke.sk, lubos.ovsenik@tuke.sk

***Abstrakt* — Free Space Optics (FSO) technológia s veľkou šírkou pásma sa stala bezdrôtovou alternatívou k optickej sieti s optickými vláknami. Medzi hlavné výhody FSO oproti vláknovej optike je jeho rýchla inštalácia, bezpečnosť prenášaných dát a značné úspory v nákladoch. Nevýhodou je, že prijatý signál v FSO prijímacom systéme je premenlivý a ťažko predikovaný. Je to hlavne spôsobené vplyvom počasia na prenosovú FSO linku. Tento článok sa zaoberá zálohovaním FSO linky RF linkou a vlastnosťami tejto hybridnej FSO/RF linky, ktorá zabezpečí vysokú dostupnosť a nepretržitú prevádzku.**

***Kľúčové slová* — FSO system, hybridný FSO/RF systém, vlastnosti hybridného FSO/RF systému, MIMO prenosový systém**

I. Úvod

FSO komunikácia, tiež známa ako bezdrôtová optická komunikácia, je širokopásmová technológia využívajúca šírenie optického lúča voľným prostredím k prenosu informácii. Point-to-point FSO prenosový systém môže dosiahnuť dátovú rýchlosť porovnateľnú s dátovou rýchlosťou káblových optických komunikačných systémov bez toho, aby došlo k zvýšeniu ceny a času na inštaláciu v porovnaní s vláknovou optikou. Avšak široké využitie FSO komunikačných systémov bolo sťažené vzhľadom na problémy spoľahlivosti. Spoľahlivosť FSO komunikačnej linky výrazne klesá v prítomnosti dymu a hmly. V dôsledku toho sa nedosiahne požadovaná dostupnosť 99,999%. Preto sa v poslednej dobe začalo čoraz intenzívnejšie pracovať na vyriešení tohto problému pomocou použitia hybridného FSO/RF komunikačného systému. Hybridný kanál môže využívať výhody oboch FSO a RF liniek [1-3]. Hybrid FSO/RF, tj. FSO systém zálohovaný RF systémom, využíva silné vlastnosti každého systému na prekonanie ich slabých stránok. Tým môže tento hybridný FSO/RF systém dosiahnuť vyššiu dostupnosť prenosovej linky, ako je dosiahnutá samostatnou FSO linkou alebo RF linkou a to preto, že RF časť hybridného FSO/RF môže zmierniť vplyv hustej hmly a FSO časť môže zmierniť vplyv dažďových zrážok. Hmla je najdôležitejším nepriaznivým faktorom FSO spojenia a dážď je najdôležitejším nepriaznivým faktorom RF spojenia.

II. Charakteristika FSO a RF Liniek

A. FSO Linka

Zo všetkých nepriaznivých vplyvov na FSO linku má najväčší vplyv hmla, ktorá môže zapríčiniť úplne zlyhanie FSO linky [1], [7]. V metrológii sa intenzita hmly vyjadruje viditeľnosťou. Uvedieme niektoré príklady útlmu a viditeľnosti pre FSO komunikačný systém s vlnovou dĺžkou 850 nm. Ľahká hmla, ktorá predstavuje viditeľnosť okolo 10 km spôsobuje útlm FSO linky 1 až 4 dB/km. Ťažká hmla, ktorá predstavuje viditeľnosť menšiu ako 0,5 km spôsobuje veľmi veľký útlm FSO linky, ktorý je väčší ako 40 dB/km a môže dosahovať hodnoty útlmu až do 300 dB/km. Útlm spôsobený hmlou závisí aj od vlnovej dĺžky prenášaného signálu. Je nižší na dlhších vlnových dĺžkach, ale rádovo je podobný.

Útlm spôsobený silným dažďom je nezanedbateľný a závisí od intenzity zrážok. Veľmi silné zrážky, čo je 90 až 100 mm/h, vyskytujúce sa v miernom klimatickom podnebí spôsobujú útlm okolo 20 dB/km.

B. RF Linka

Hmla spôsobuje do istej miery straty v RF prenosovom kanáli, ale tieto straty na rozdiel od strát v FSO prenosovom kanáli sú zanedbateľné [3], [7]. Tieto straty spôsobené hmlou závisia

od obsahu kvapôčok vody vo vzduchu a teploty vzduchu. V porovnaní s FSO, pre mm vlny je hlavnou príčinou výpadku spojenia silný dážď. Útlm RF linky môže byť 30-35 dB/km pri zrážkach 90 mm/h a je iba slabo závislý od prenosovej frekvencie.

Záložná rádiová linka teda kompenzuje výpadok FSO linky spôsobený negatívnym účinkom hmly. So záložnými RF linkami sa počíta už od prvých dní FSO. Z toho hľadiska je významný rozdiel medzi RF linkami v minulosti a dnešnej dobe. RF prenosové linky s prenosovými rýchlosťami Gbit/s v minulosti neexistovali. Preto všetky publikácie v tejto oblasti vrátane [7], [8] uvažujú RF záložnú linku pre zvýšenie dostupnosti v hmlistom období pri oveľa nižšej rýchlosti prenosu dát ako pri peknom a jasnom počasí. Avšak priemerné zníženie rýchlosti by nemalo byť menšie ako 10 násobok. Z hľadiska dostupnosti najvhodnejšie záložné linky by boli linky pracujúce s frekvenciu pod 10 GHz (napr. [8]), kde vplyv hmly a dažďa na prenášaný signál je malý. V skutočnosti však toto riešenie je veľmi ťažko dosiahnuteľné z rôznych dôvodov. Rýchlosť prenosu záložnej linky by musela byť aspoň 100 Mbit/s a neskôr možno 300 – 1000 Mbit/s, ktorá je nad limit aktuálnych technických možností. Jednou z výhod FSO je bezlicenčné pásmo, zatiaľ čo táto výhoda pre navrhovane zálohovacie RF linky s frekvenciami pod 10 GHz je ťažko dosiahnuteľná (i keď to neplatí pre 2,4 a 5 GHz ISM/UNI pásma, ale preťaženie v týchto pásmach je extrémne veľké).

Zálohovanie mm vlnami môže byť navrhované a riešené v rôznych frekvenčných pásmach. Náš model uvažuje RF linku, ktorá je tvorená 60 GHz anténou kvôli jej špecifickým vlastnostiam, ktoré sú presne určené a navrhnuté v nasledujúcej kapitole. 60 GHz linka je v podstate zdarma (bez licencie), taktiež je bez vnútro systémovej interferencie a šírka pásma je dostatočne vysoká aj pre Gb/s prenosy. Na druhej strane toto frekvenčné pásmo možno použiť iba na krátke spojenie, typicky do 1 km.

C. Hybridná FSO/RF linka

Hybridnú FSO/RF prenosovú linku uvažujeme z dôvodu zálohovania primárnej FSO prenosovej linky RF linkou, ktorá môže zlyhať v dôsledku nepriaznivých vplyvov, hlavne počas hustej hmly. Tento hybridný systém pozostávajúci z dvoch liniek pracuje paralelne a súbežne. To znamená, že obidvoma linkami sa šíri ten istý signál v rovnakom časovom okamihu. Podrobný popis poveternostných vplyvov je uvedený v [7]. Pôsobenie turbulentných vplyvov na optické lúče šíriace sa vo voľnom prostredí a meranie chýb (P*ointing Errors*) je popísaný v [2], [6], [9]. Šíriace charakteristiky mm vĺn pre pozemné a satelitné komunikácie sú popísané v [1], [3], [4]. Konkrétnymi charakteristikami hybridného FSO/MIMO RF systému sa budeme zaoberať v nasledujúcej kapitole.

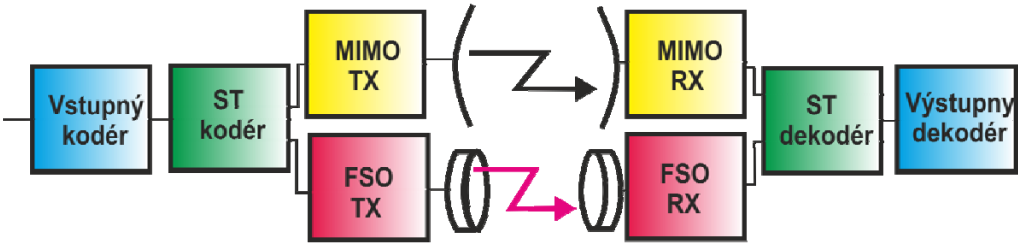
III. TECHNICKÉ ASPEKTY PARALELNEJ FSO/RF KOMUNIKAČNEJ LINKY

Ak FSO linka pracuje paralelne spolu s RF linkou, tento systém budeme považovať za osobitný prípad MIMO systému. Jeho odlišnosť od tradičného MIMO je v tom, že prijímacie antény nevidia signál z paralelnej FSO linky. Diagonálna kanálová matica tohto hybridného systému je zobrazená v (1) ako:

$$H = \begin{pmatrix} h_1 & 0 & \cdots & 0 & 0 \\ 0 & h_2 & \cdots & 0 & 0 \\ 0 & 0 & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & h_i \end{pmatrix} \tag{1}$$

kde $h_i \quad i = 1, 2, 3, \ldots$ je prenosový koeficient súvisiaci s útlmom signálu (*z angl. related to fading*).

Uvažovaný paralelný systém je zložený z jednej časti, ktorú tvorí FSO linka a druhej časti ktorú tvorí MIMO prenosový systém (*Obr. 1*). Avšak ak tieto dve časti uvažujeme ako jeden celkový prenosový systém, potom môžeme použiť priestorovo-časové (*Space Time (ST)*) trelisové kódovanie. ST kodér a dekodér slúžia ako vnútorný kodek (napr. LDPC). Úlohou ST kódovania je zvýšiť kódovací zisk a to s kódovou rýchlosťou 1. Pripomeňme, že tu nebudeme riešiť kódovaciu stránku, ale niektoré podrobnosti o ST kódov optimalizovaných pre paralelný kanál MIMO nájdeme v [5].



Obr. 1 Bloková schéma paralelného FSO/MIMO RF systému s ST kódovaním.

V našom prípade nás zaujíma kapacita kanálu. Keďže tento kanál nie je ergodický (kanál je v skutočnosti konštantný počas prenesenia tisícky alebo miliónov symbolov) je pre nás zaujíma chybová kapacita v bit/s/Hz (*Outage Capacity*). V tomto príspevku budem túto kapacitu označovať ako Outage Kapacita a Outage Pravdepodobnosť. Kanálovú kapacitu určíme zo štandardnej MIMO kapacitnej formuly uvedenú nižšie pre deterministický prípad:

$$C = \log_2\left[\det\left(I_m + \frac{\rho}{n} H * H^+\right)\right] = \log_2 \prod_{i=1}^{n}\left(1 + \frac{\rho}{n}|h_i|^2\right) \tag{2}$$

kde ρ je pomer signálu k šumu (SNR),
n je počet vysielacích antén,
m je počet prijímacích antén,
I_m je jednotková matica rádu m a index + označuje Hermitovský znak.

Ak chceme použiť tento vzťah na určenie kapacity pre hybridný FSO/RF systém, potrebujeme vykonať nasledujúce modifikácie: Musíme poznamenať, že v našom prípade ρ je elektrická veličina. V optickej DD je ρ priamo úmerný optickému výkonu umocnenému na druhú. V optickej heterodynnej detekcii je SNR úmerný optickému výkonu, kým šumový výkon, v tomto prípade, je výstrelový šum lokálneho oscilátora. V našom prípade budeme uvažovať SNR pre FSO linku a SNR pre RF linku. SNR pre RF linku označíme indexom ρ_2 a SNR pre FSO linku označíme indexom ρ_1.

Optický SNR v DD môžeme určiť pomocou prijatej optickej energie, ktorá je približne daná ako [9]:

$$P_{r,opt} = P_{t,opt} \cdot \left(\frac{\Phi}{D\nu}\right)^2 \tag{3}$$

kde $P_{t,opt}$ je vyslaná optická energia,
Φ je priemer prijímacieho objektívu,
D je vzdialenosť prijímača a vysielača,
ν je uhol rozptylu šíriaceho lúča.

Potom elektrický SNR pre FSO linku je:

$$\rho_1{}^2 = \left(\frac{i_{signál}}{i_{šum}}\right)^2 = \left(\frac{R_d P_{r,opt}}{i_{šum}}\right)^2 \tag{4}$$

kde $i_{šum}$ je šumový prúd elektrického zosilňovača (predovšetkým tepelný) daný výrobcom a R_d je spoľahlivosť fotodetektora. Aplikáciou vzorcov (2), (3), (4) dostaneme vzťah na určenie kapacity pre DD systém:

$$C = \log_2\left(\left(1 + \rho_1{}^2 |h_1|^4\right) \times \left(\prod_{i=2}^{n}\left(1 + \frac{\rho_2}{n}|h_i|^2\right)\right)\right) \tag{5}$$

A. Správanie sa systému v prípade silného dažďa

Všimnime si, že dážď je zvyčajne opísaný fenoménom zlyhania (výpadku) linky pre oba šíriace média – optiku a mm vlny –, ktoré sú naň citlivé na rôznej úrovni. Podľa [7], útlm zapríčinený dažďom pre mm vlny a FSO linku môžeme vyjadriť vzťahom:

$$A = \chi \cdot D = kDR^{\alpha} \quad dB \tag{6}$$

χ sú straty na km,
R je hustota zrážok v mm/hodinu,
D je efektívna vzdialenosť medzi vysielačom a prijímačom v km,
k a α sú frekvenčne a polarizačne závislé parametre.

Pre 60 GHz frekvenčné pásmo platí:

$$k_{horiz} = 0{,}955\,; \quad k_{vert} = 0{,}888\,; \quad \alpha_{horiz} = \alpha_{vert} = 0{,}772 \tag{7}$$

V optickom pásme sú k a α vyjadrené ako:

$$k = 1{,}076\,; \qquad \alpha = 0{,}67 \tag{8}$$

Hodnoty útlmu prenosovej linky závisle od dažďa sú uvedené v *Tab. 1* [5].

Tabuľka 1
HODNOTY ÚTLMU SPÔSOBENÉHO MNOŽSTVOM ZRÁŽOK PRE FSO A RF LINKU

Pravdepodobnosť	10^{-4}	10^{-5}
R, mm/h	35.9	96.2
χ_1, dB/km	11.75	22.94
χ_2, dB/km	15.15	32.42

$|h_i|^2$ hodnoty sú dané ako:

$$|h_i|^2 = \frac{10^{-\chi_i \frac{D}{10}}}{D^2}\,; \; i = 2,3,...,m \tag{9}$$

B. Správanie sa systému v prípade hmly

Musíme poznamenať, že vplyv hmly na FSO linku je veľmi nepriaznivý pre dostupnosť a spoľahlivosť linky. Ale pre celý paralelný systém hmla nemá taký škodlivý vplyv ako iba na FSO prenosovú linku. Predpokladajme, že do roka je 16 % dní hmlových a hmla trvá v priemere 6 hodín do dňa [1]. S rovnakou pravdepodobnosťou môžeme povedať, že sa jedná o veľmi ľahkú hmlu s útlmom FSO linky 4 dB/km, ľahkú hmlu s útlmom FSO linky $\leq$ 20 dB/km a ťažkú hmlu s útlmom FSO linky >120 dB/km. Útlm spôsobený hmlou pre RF linku s frekvenčným pásmom 60 GHz pre tie isté situácie hmly je 0 dB/km; 0,25 dB/km a 3 dB/km. Spoločný výskyt hmly a silného dažďa je nepravdepodobný a môžeme ho vylúčiť. Preto za predpokladu vyššie uvedeného útlmu hmlou RF zariadenie musí zabezpečiť dostupnosť linky pre svoj kanál okolo 1 % z času, 1 % pre spoločný RF/FSO silne a 1% pre slabo redukovanú kvalitu FSO linky.

IV. ŠPECIFIKÁCIA FSO/MIMO RF SYSTÉMU

Hlavnou špecifikáciou paralelnej FSO/MIMO linky je K, ktoré označuje počet bitov na kanál. Táto forma kapacity je viac využívaná v diverzitných systémoch než (bit/s)/Hz. Totižto jeden kanál je formovaný celým diverzitným systémom, či už zaberá jedno alebo viac ako jedno frekvenčné pásmo. Pri voľbe K berieme do úvahy kódový pomer výstupného kódera (*Obr. 1*). Rýchlosť ST kódu je 1. Definíciu *Outage* Pravdepodobnosti popíšeme ako :

$$P_{out} = \Pr\{C < K\} \tag{10}$$

Ak C je dané ako funkcia SNR (signál/šum), vzťah (10) môžeme vyjadriť pomocou komutatívnej distribúcie $C^{-1}(SNR_{C=K})$.

$$P_{out} = \Pr\left\{SNR < C^{-1}(SNR_{C=K})\right\} \tag{11}$$

V princípe z každého rušivého fyzikálneho javu môže byť vytvorená joint distribúcia. Avšak vzhľadom na to, že dažďové a hmlové periódy nie sú stacionárne, tento prístup je potom formulovaný bez nejakého veľkého významu. Namiesto toho Outage Pravdepodobnosť môžeme vyjadriť ako:

$$P_{out} = \Pr\{clear\,air\} \cdot \Pr\{out\,|\,clear\,air\} + \Pr\{fog\} \cdot \Pr\{out\,|\,fog\} + \Pr\{out\,|\,rain\} \tag{12}$$

Prvý a druhý príspevok vo vzťahu (12) pre RF linku sú veľmi malé. Preto systémový návrh hybridnej linky môže byť založený len na treťom príspevku. Podľa FSO systému budeme

pracovať s binárnou moduláciou a s optickou priamou detekciou a to OOK pri FSO a BPSK pri RF.

A. Matematické vyjadrenie Outage Kapacity pre FSO/MIMO linku

Teraz si uvedieme hodnoty výpočtu Outage Kapacity pre FSO systém, RF systém a pralélny systém podľa odvodených vzťahov. Vplyv nepriaznivých fenoménov budeme počítať pre prototyp systému, ktorý bude využívať technické parameter systémov, ktoré máme aj v našom laboratóriu (LOS KEMT FEI TUKE). Parametre FSO linky sú špecifikované v *Tab. 2*. Pre RF linku budeme predpokladať priemerný *Fade Margin* 45 dB (pre linkovú vzdialenosť 1 km).

Tabuľka 2
PARAMETRE FSO SYSTÉMU

Vlnová dĺžka	1550 nm
Vysielací výkon	40 mW
Uhol divergencie	2 mrad
Úroveň modulácie	Dvojková alebo štvorková
Priemer prijímača	25 cm
Citlivosť	1 A/W (t.j. η_q=8)
Optická detekcia	DD alebo heterodýna
Šumový prúd prijímača	200 nA
Referenčná dĺžka prenosovej linky	1 km (ρ_1=35 dB v DD 65 dB v heterodynej)

Maximálna prenosová vzdialenosť je určená pre Outage Pravdepodobnosť10^{-4} a 10^{-5} pre rôzne linky a to: iba pre FSO linku, iba pre RF linku a pre hybridnú linku. Hodnoty sú počítané pre DD binárnu moduláciu. V *Tab.3* sú uvedené hodnoty ktoré boli vypočítané pomocou vyššie uvedených vzťahov. Vzorový výpočet pre jednu hodnotu z tabuľky je uvedený pod tabuľkou.

Tabuľka 3
Outage kapacita v závislosti od prenosovej dĺžky; DD binárna modulácia $P_{out}=10^{-5}$; ρ_1=35 dB; ρ_2=45 dB

D/km	$C_{out/RF}$	$C_{out/FSO}$	$C_{out/Hybrid}$	Hraničná dĺžka pre
0,5	11,6	19,6	31,2	
0,7	8,4	14,6	23	
0,9	5,6	10,1	15,7	
1	4,3	8	12,3	
1,15	2,5	5	7,5	
1,25	1,5	3	4,5	
1,35	0,8	1,55	2,35	
1,36	0,7	1,3	2	
1,4	0,56	1	1,56	RF
1,45	0,38	0,59	0,97	FSO
1,5	0,25	0,33	0,58	HYBRID
1,55	0,17	0,18	0,35	
1,6	0,11	0,09	0,2	

Vzorový výpočet pre štvrtý riadok v tabuľke:

$D = 1km$; $\rho_1 = 35\ dB$; $\rho_2 = 45\ dB$; χ_1 a χ_2 sú z *Tab.1*

Podľa vzťahu (9) vypočítame hodnoty $|h_1|^2$ a $|h_2|^2$ a to:

$$|h_1|^2 = \frac{10^{-\chi_1 \cdot \frac{1}{10}}}{1^2} = \frac{10^{-22,94 \cdot \frac{1}{10}}}{1^2} = 10^{-2,294} = 5,08 \cdot 10^{-3} \tag{13}$$

$$|h_2|^2 = \frac{10^{-\chi_2 \cdot \frac{1}{10}}}{1^2} = \frac{10^{-32,4 \cdot \frac{1}{10}}}{1^2} = 10^{-3,24} = 5,75 \cdot 10^{-4} \tag{14}$$

Outage Kapacitu hybridnej linky určíme pomocou vzťahu (5):

$$\begin{aligned} C_{out/Hybrid} &= \log_2\left(\left(1+\left(10^{3,5}\right)^2 \cdot \left(5,08 \cdot 10^{-3}\right)^2\right) \cdot \left(1+\left(10^{4.5} \cdot 5,75 \cdot 10^{-4}\right)\right)\right) = \\ &= \log_2 4969,625 = \underline{12,3\ bit/s/\ Hz} \end{aligned} \tag{15}$$

$$C_{out/FSO} = \log_2\left(1 + \left(10^{3,5}\right)^2 \cdot \left(5{,}08 \cdot 10^{-3}\right)^2\right) = \log_2 259{,}06 = \underline{8\ bit/s/Hz} \quad (16)$$

$$C_{out/RF} = \log_2\left(1 + \left(10^{4.5} \cdot 5{,}75 \cdot 10^{-4}\right)\right) = \log_2 19{,}183 = \underline{4{,}3\ bit/s/Hz} \quad (17)$$

V Tab.4 sú vypočítané hodnoty Outage Kapacity s Outage Pravdepodobnosťou 10^{-4}.

Tabuľka 4
Outage kapacita v závislosti od prenosovej dĺžky; DD binárna modulácia $P_{out}=10^{-4}$; ρ_1=45 dB; ρ_2=35 dB

D/km	$C_{out/RF}$	$C_{out/FSO}$	$C_{out/HYBRID}$	Hraničná dĺžka pre
1	10	15.4	25.4	
1,5	6,3	9,2	15,5	
2	3	3,8	6,8	
2,25	1,8	1,6	3,8	
2,4	1,2	0,76	1,96	FSO
2,5	0,87	0,43	1,3	
2,6	0,62	0,22	0,84	RF
2,7	0,44	0,12	0,56	HYBRID
2,8	0,3	0,06	0,36	

V. ZÁVER

Zisteným dôsledkom týkajúcim sa FSO samotnej linky je, že vplyv dažďa na hybridnú paralelnú linku viac spôsobuje nedostupnosť tejto hybridnej linky ako vplyv hmly. Dážď je iba vplyv životného prostredia, na ktorý sú tieto paralelné linky citlivé (vplýva na určenie dostupnosti alebo nepriamo na vzdialenosť prenosovej linky) nedbajúc na ďalšie vplyvy pôsobiace na samotnú FSO linku. Ak zoberieme do úvahy výsledky kapitoly IV, môžeme tvrdiť, že ak RF linka s mm vlnami pracuje paralelne s FSO linkou, tak FSO linka slúži skôr ako záložná linka pre RF linku.

Ak uznáme, že táto paralelná linka môže byť považovaná za špeciálny prípad takzvaného paralelného MIMO systému, kvalita prenosu môže byť zlepšená aplikovaním priestorovo-časového trelisového kódovania (*Space-Time Trellis coding (ST)*) okrem obvyklých kanálových kódovaní. ST kódovanie s rýchlosťou kódu 1 môže priniesť dodatočný kódovací zisk.

POĎAKOVANIE

Podporujeme výskumné aktivity na Slovensku/ Projekt je spolufinancovaný zo zdrojov EÚ. Tento článok bol vypracovaný v rámci projektu "Centrum excelentnosti integrovaného výskumu a využitia progresívnych materiálov a technológií v oblasti automobilovej elektroniky", ITMS 26220120055 (50%). Táto práca bola taktiež podporovaná a realizovaná z projektu Rozvoj Centra informačných a komunikačných technológií pre znalostné systémy (ITMS kód: 26220120030) na základe podpory operačného programu Výskum a vývoj financovaného z Európskeho fondu regionálneho rozvoja (50%).

LITERATÚRA

[1] B. Arbesser-Rastburg, "ESA Publication, Final Report", *COST Action 255*, Nordwijk, The Netherland, 2002.

[2] D. K. Borah, D. G. Voelz, "Pointing error effects on communication links", *Journal Lightw.Techn*, vol. 27, no 18, September 2009, pp. 3965-3973.

[3] L. Castanet, „Influence of the Variability of the Propagation Channel on Mobile", *Fixed Multimedia and Optical Satellite Communications*, Aachen, Germany, Shaker Verlag, 2008.

[4] M. Hall, ed., „Final Report COST Action 235", *European Commission Publishing,* Luxambourg, 1996.

[5] P. Horváth, I. Frigyes, „Application of the MIMO concept in millimeter-wave broadband wireless access networks", *Int. J. of Wireless Information Networks,* vol. 11, no. 4, , Október 2004, pp.217-225.

[6] G. K. Karagiannidis & al., „Optical wireless communication with heterodyne detection...", *Journal Lightw.Techn*, vol. 27, no 20, Október 2009, pp. 4440-4445.

[7] E. Leitgeb, & al., „Weather effects on hybrid FSO/RF communication link,…", *IEEE J.Select. Areas in Comm.*, vol. 27, no. 9, December 2009, pp. 1687-1697.

[8] R. Luna, D. K. Borah, H. Tapse, „Behavior of hybrid optical/RF channels...", *Proc. IEEE GLOBECOM*, Honolulu, HA, November-December 2009.

[9] A. Prokes, „Modeling of atmospheric turbulence effect on terrestrial FSO links", *Radioengineering*, vol 18, no 1, April 2009, pp.42-47.

Vlastnosti Hybridnej FSO/RF linky so 60 GHz RF záložnou linkou

Pavol MIŠENČÍK, Ľuboš OVSENÍK, Ján TURÁN

Katedra Elektroniky a multimediálnych telekomunikácií, Fakulta Elektrotechniky a Informatiky, Technická Univerzita v Košiciach, Slovenská Republika

pavol.misencik@tuke.sk, lubos.ovsenik@tuke.sk, jan.turan@tuke.sk

***Abstrakt* — Free Space Optics (FSO) prenosové linky ponúkajú gigabitové prenosové rýchlosti dát a nízku zložitosť systému, ale trpia vplyvom vlastností atmosféry. Rádiofrekvenčné linky majú nižšie prenosové rýchlosti, ale sú relatívne odolné voči poveternostným podmienkam. Hybridná linka môže kombinovať výhody oboch liniek a tým dosiahnuť vyššiu dostupnosť spojenia. Publikácia sa bude zaoberať vplyvmi atmosféry na FSO linku a 60 GHz RF linku. Rozoberieme tiež vlastnosti a výhody použitia 60 GHz RF linky oproti ostatným RF linkám.**

***Kľúčové slová* — FSO system, hybridný FSO/RF systém, vlastnosti hybridného FSO/RF systému, MIMO prenosový systém**

I. Úvod

Hybridná FSO/RF komunikačná linka pozostáva z dvoch hlavných systémov. Prvým primárnym komunikačným systémom je FSO systém, ktorý zabezpečuje primárnu FSO optickú linku. Druhým zálohovým systémom je RF systém, ktorý zabezpečuje dostupnosť hybridnej linky prípade zlyhania primárnej FSO komunikačnej linky. Treťou časťou je prepínač, ktorý vykonáva prepínanie medzi primárnou a záložnou linkou a taktiež zabezpečuje spätné prepnutie liniek pri obnovení spojenia primárnej FSO linky. V tomto dokumente sa nebudeme zaoberať prepínaním medzi týmito dvoma linkami. V prvej časti tohto dokumentu sa budeme venovať výhodám a problémom FSO komunikačnej linky. V druhej časti dokumentu sa budeme venovať výhodám 60 GHz RF linky, a taktiež vplyvom atmosféry na túto RF komunikačnú linku.

II. Charakteristika FSO Linky

FSO linku, respektíve optickú komunikáciu voľným prostredím, je možné definovať ako telekomunikačnú technológiu, ktorá využíva šírenie svetla na prenos informácií medzi dvoma bodmi. Ide o širokopásmovú komunikačnú technológiu na priamu viditeľnosť, tzv. „Line-Of-sight“ (LOS) technológiu, ktorá používa optické pulzne modulované signály na bezdrôtový prenos dát. Oproti vláknovej optike, kde sa svetelné impulzy prenášajú skleneným vláknom, sa impulzy svetla prenášajú úzkym zväzkom lúča prostredníctvom atmosféry, teda voľným prenosovým prostredím. FSO technológia teda na tento prenos využíva svetelné zväzky namiesto rádiových vĺn, ako je to pri bežných bezdrôtových technológiách.

Systém FSO je založený na konektivite medzi optickými bezdrôtovými jednotkami. Základom FSO systému je FSO optická bezdrôtová jednotka pozostávajúca z optického transceivera (vysielač/prijímač), ktorý poskytuje plne duplexný prenos.

A. Frekvenčná oblasť využitia FSO systému

Systémy FSO využívajú pre prácu infračervenú oblasť svetla (IR), ktoré susedí so spektrom viditeľného svetla (*Obr. 1*), čo znamená, že pre ľudské oko sú optické zväzky neviditeľné [1]. V tomto dôsledku je veľmi dôležitá bezpečnosť pri používaní týchto FSO systémov.

Pracovná oblasť FSO systému vo frekvenčnej oblasti spadá do rozsahu stoviek terahertzov (1 THz = 10^{12} Hz), čomu analogicky zodpovedajú prevádzkové vlnové dĺžky, rádovo v desatinách až jednotkách mikrometrov. Vlnová dĺžka závisí od frekvencie vlnenia a rýchlosti šírenia vlny. Vzťah, ktorý môžeme použiť na prevod frekvencie a vlnovej dĺžky je:

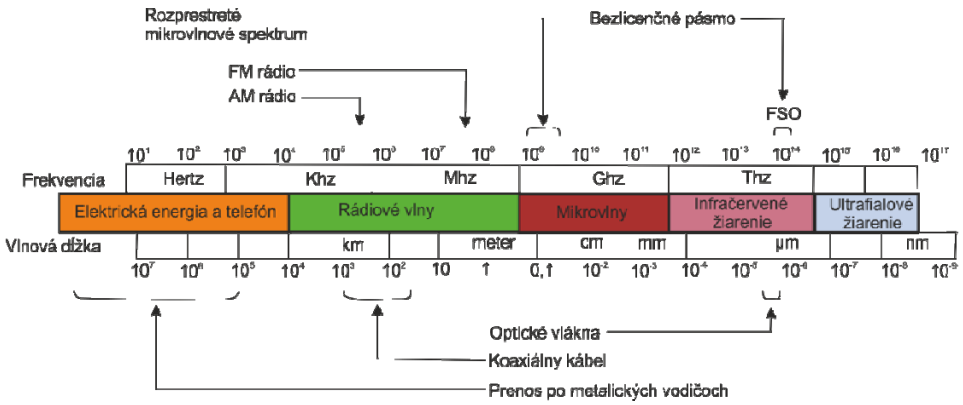
$$\lambda = \frac{v}{f} \tag{1}$$

kde

λ = vlnová dĺžka v m,
ν = rýchlosť šírenia vlny v m/s,
f = frekvencia vlnenia v Hz.

Tento vzťah sa dá vyjadriť aj zjednodušene (približne) v tvare

$$\text{vlnová dĺžka } \lambda[m] = 300 / \textit{frekvencia } [MHz] \tag{2}$$



Obr. 1 Frekvenčné spektrum.

B. Využitie FSO systému

Oblasť využitia FSO systému je pre rôzne aplikácie veľmi široká. Pohodlne môže vďaka veľkej šírke pásma dopĺňať aj technológie ako sú optické káblové a rádiové komunikácie, optické vláknové komunikačné technológie a pod. Väčšina koncových užívateľov sú v krátkej vzdialenosti od chrbticovej siete (1km a menej), čo je veľkou výzvou pre aplikáciu FSO do tejto oblasti. Ďalšie oblasti vhodné na využívanie FSO systému sú:

Last mile access - FSO vypĺňa medzeru šírky pásma, ktorá existuje na mieste, kde sa končí vedenie optických vlákien a signál je potrebné preniesť ku koncovým používateľom. FSO tak vytvára priamo ideálny spôsob transportu signálu vzhľadom na vzdialenosť (50 m až niekoľko kilometrov) s prenosovou rýchlosťou (1 Mbps až 2,5 Gbps) [5].

Záložná linka pre optickú sieť - používa sa ako záložný komunikačný zdroj proti strate informácií alebo komunikačného kolapsu v prípade poškodenia alebo nedostupnosti hlavnej optickej vláknovej linky.

Záloha pri katastrofách - keď je potrebné uvažovať aj s dočasnou resp. náhradnou linkou (potrebná neustála komunikácia) v prípade, že dôjde k neočakávanému kolapsu hlavnej siete.

Komunikačná sieť v areáli – využívanie FSO na prepojenie siete v rôznych areáloch.

Komplikovaný terén - rieky, rušné ulice, vlakové trate alebo miesta, kde je príliš drahá inštalácia optického kábla, sa FSO stáva alternatívou premostenia týchto bariér.

C. Vplyv atmosféry na FSO linku

Vzhľadom na FSO systémy a vlnové dĺžky používané týmito systémami sú najdôležitejšie častice hmly. Hmla je zložená z jemných čiastočiek vody (0.5 µm – 2 µm) rozptýlených vo vzduchu nízko nad povrchom Zeme. Molekuly hmly sú porovnateľne veľké s vlnovou dĺžkou používaného optického žiarenia. Tieto jemné čiastočky rozptyľujú svetlo a tým znižujú viditeľnosť nízko nad povrchom. S nárastom hmly stúpa aj vlhkosť vzduchu. Pri poklese viditeľnosti pod 1 km, je vzduch skoro plne nasýtený (95 %).

Preto je vplyv počasia a najmä hmly veľmi kritický pre FSO systémy. Napríklad v hustej hmle môže tlmenie narásť na hodnotu 300 dB/km (toto číslo závisí od miestnych klimatických pomerov, pri morskej hmle môže tlmenie dosiahnuť hodnotu až 450 dB/km), zatiaľ čo pre RF milimetrové vlny môže byť tlmenie pri rovnakých podmienkach iba okolo 10 dB/km. Takéto veľké tlmenie je limitujúce pretože kvôli bezpečnosti sa musia používať malé vysielacie výkony FSO systémov.

Hmla modifikuje prenášané svetlo, čím môže dôjsť k strate informácii. Tlmenie zapríčinené hmlou výrazne skracuje použiteľné vzdialenosti jednotlivých FSO liniek. Vplyv dažďa nie je taký veľký, pretože kvapky dažďa majú zvyčajne veľkosť 0,2 mm [6]. Vlnové dĺžky, ktoré sa využívajú pri tomto prenose sú v rozmedzí nm, preto kvapky dažďa spôsobujú podstatne nižší útlm ako hmla. Rozptyl spôsobený zrážkami fotónov s kvapkami dažďa sa nazýva neselektívny

rozptyl, keďže veľkosť kvapky je oveľa väčšia než prenášaná vlnová dĺžka. Tlmenie spôsobené dažďom môžeme určiť pomocou vzťahu:

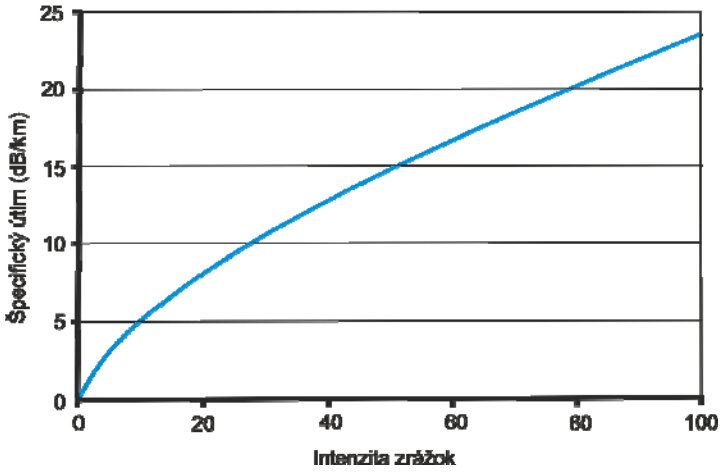
$$\alpha_{rain} = 1.076 R^{\frac{2}{3}} [dB/km] \quad (3)$$

kde R predstavuje množstvo zrážok v mm/hod.

Hodnota tlmenia sa mení s dennou hodinou a s viditeľnosťou. Pri zníženej viditeľnosti, ktorej príčina je hmla, výrazne stúpa tlmenie FSO systému. Špecifický útlm (dB / km) spôsobený dažďom je všeobecne opísaný vzťahom:

$$\gamma_{rain} = k.R^{\alpha} \quad (4)$$

Obr. 2 zobrazuje typické varianty konkrétnych útlmov (dB / km) v dôsledku zrážok [6].



Obr. 2 Závislosť špecifického útlmu od zrážok.

Typické hodnoty tlmenia a dosahu FSO systému (zisk vysielača ~20 dB) sú uvedené *Tab.1* [11].

Tabuľka 1
VPLYV HMLY A DAŽĎA

Poveternostné Podmienky	Zrážky		Viditeľnosť [m]	Straty [dB/km]	Typický dosah [m]
Hustá hmla			0 50	-271,65	122
Silná hmla			200	-59,57	490
Stredná hmla	Sneh		500	-20,99	1087
Ľahká hmla	Sneh	Prietrž mračien	700 1000	-12,65 -9,26	1565 1493
Riedka hmla	Sneh	Silný dážď	1900 2000	-4,22 -3,96	3238 3369

III. CHARAKTERISTIKA RF LINKY

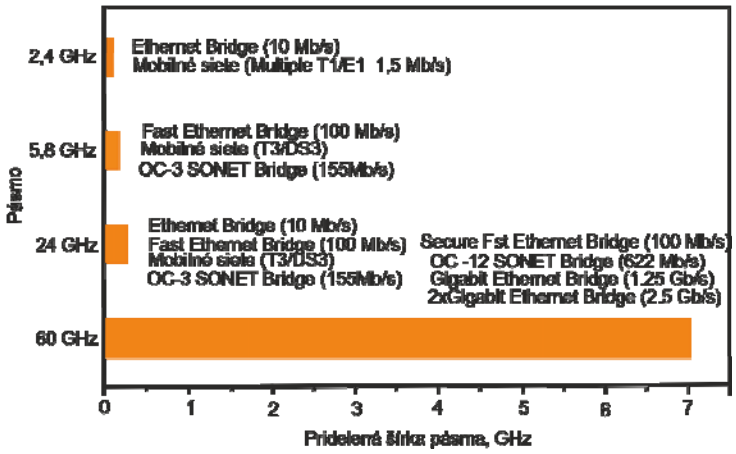
Hmla spôsobuje do istej miery straty v RF prenosovom kanáli, ale tieto straty sú na rozdiel od strát v FSO prenosovom kanáli zanedbateľné. Straty spôsobené hmlou závisia od obsahu kvapôčok vody vo vzduchu a teploty vzduchu. V porovnaní s FSO, pre mm vlny je hlavnou príčinou výpadku spojenia silný dážď. Útlm RF linky môže byť 30-35 dB/km pri zrážkach 90 mm/h a je iba slabo závislý od prenosovej frekvencie.

RF rádiová linka v hybridnej FSO/RF linke kompenzuje výpadok FSO linky spôsobený negatívnym účinkom hmly. So záložnými RF linkami sa počíta už od prvých dní FSO. Z toho hľadiska je významný rozdiel medzi RF linkami v minulosti a dnešnej dobe. RF prenosové linky s prenosovými rýchlosťami Gbit/s v minulosti neexistovali. Preto všetky publikácie v tejto oblasti vrátane [3,4] uvažujú RF záložnú linku pre zvýšenie dostupnosti v hmlistom období pri oveľa nižšej rýchlosti prenosu dát ako pri peknom a jasnom počasí. Avšak priemerné zníženie rýchlosti by malo byť menšie ako 10 násobok. Z hľadiska dostupnosti najvhodnejšie záložné linky by boli linky pracujúce s frekvenciu pod 10 GHz (napr. [4]), kde vplyv hmly a dažďa na prenášaný signál je malý. V skutočnosti však toto riešenie je veľmi

ťažko dosiahnuteľné z rôznych dôvodov. Rýchlosť prenosu záložnej linky by musela byť aspoň 100 Mbit/s a neskôr možno 300 – 1000 Mbit/s, ktorá je nad limit aktuálnych technických možností. Jednou z výhod FSO je bezlicenčné pásmo, zatiaľ čo táto výhoda pre navrhovane zálohovacie RF linky s frekvenciami pod 10 GHz je ťažko dosiahnuteľná (aj keď to neplatí pre 2,4 a 5 GHz ISM/UNI pásma, ale preťaženie v týchto pásmach je extrémne veľké).

A. Vlastnosti 60 GHz RF linky

V roku 2001 federálna komunikačná komisia (FCC) zaviedla nepretržitý blok 7 gigahertzového (GHz) spektra medzi 57 a 64 GHz pre bezdrôtové komunikácie. Hlavným faktorom tohto zavedenia 7 GHz pásma, vzhľadom na obchodné podmienky, je to, že spektrum je bez licencie. Inými slovami povedané, operátor nemusí kupovať licenciu od FCC pred začatím práce zariadenia v tomto frekvenčnom pásme. Licenčný proces je zvyčajne veľmi drahý a časovo náročný. Do tejto doby šírka pásme pre bezlicenčné pásma bola menšia ako 0,3 GHz. Na *Obr.3* môžeme vidieť porovnanie bezlicenčných pásiem a ich aplikácie.



Obr. 3 Bezlicenčné pásma a ich aplikácie.

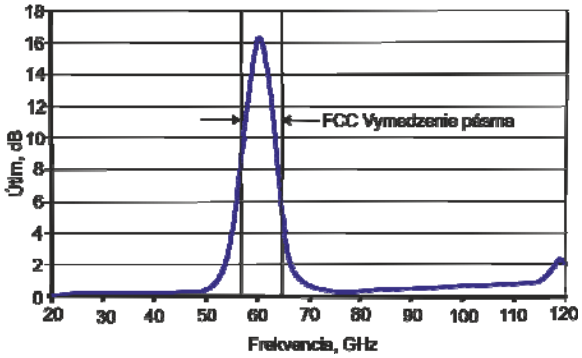
57-64 GHz pásmo sa nachádza v pásme milimetrových vĺn z časti elektromagnetického spektra, kde sa vlnová dĺžka pohybuje od desať milimetrov (30 GHz) až do jedného milimetra (300 GHz). Pre komerčné bezdrôtové aplikácie, táto časť milimetrových vĺn RF spektra bola z veľkej časti nevyužitá. V dnešnej znalosť technológií v oblasti milimetrových vĺn umožňujú obojsmernú bezdrôtovú komunikáciu s takou prenosovou rýchlosťou dát, ktorá predtým mohla byť možná len s optickým vláknom. Okrem vysokej prenosovej rýchlosti, ktorú je možné dosiahnuť v tomto frekvenčnom pásme, energetické šírenie v 60 GHz pásme má jedinečné vlastnosti, ktoré umožňujú rad ďalších výhod, ako je vynikajúca odolnosť proti rušeniu, vysoká bezpečnosť a opätovné využívanie frekvencie.

Absorpcia kyslíkom

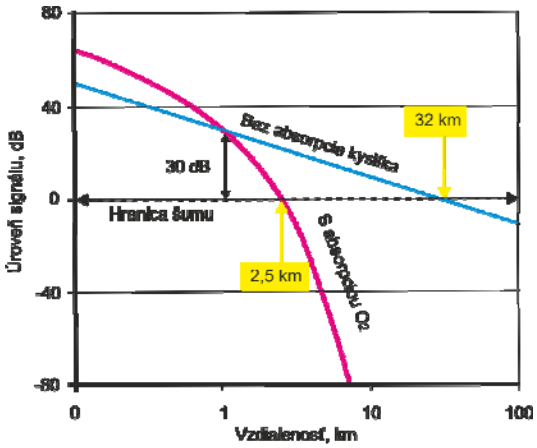
Point-to-point bezdrôtové systémy pracujúce v 60 GHz frekvenčnom pásme boli využívané spravodajskými komunitami pre vysokú bezpečnosť komunikácie a armádou pre satelitné komunikácie. Výhody tohto frekvenčného pásma závisia od absorpcie elektromagnetickej energie molekulami kyslíka (O_2) (*Obr. 4*) [6]. Táto absorpcia je oveľa výraznejšia pri frekvencii 60 GHz než u nižších frekvencií zvyčajne používaných pre bezdrôtovú komunikáciu. Absorpcia tlmí 60 GHz signály na diaľku, takže prenos signálu na dlhú vzdialenosť za prijímač nie je možný. Z toho dôvodu 60 GHz pásmo je vynikajúcou voľbou pre komunikácie na prenosovú vzdialenosť do 1 km.

Ďalším dôsledkom O_2 absorpcie je, že žiarenie z jednej 60 GHz linky nebude ovplyvňovať druhú 60 GHz linku v rovnakej geografickej oblasti. Tento útlm absorpciou umožňuje znova použitie 60 GHz frekvenčného pásma v tej istej geografickej oblasti. Pri spojeniach dlhších rozsahov (5 km a viac) to nie je možné. Ako príklad, porovnajme dve rôzne linky, jednu pracujúcu s frekvenčným pásmom 60 GHz a druhú linku pracujúcu s frekvenčným pásmom, ktoré je menej ovplyvnené absorpciou O_2. Napríklad táto linka môže pracovať tiež s bezlicenčným frekvenčným pásmom 2,4 GHz alebo 24 GHz. Uvažujme typickú situáciu, kde obe linky pracujú na vzdialenosť jedného kilometra s výstupným vysielacím výkonom 30 dB nad úrovňou šumu. Na *Obr. 5* je zobrazená klesajúca úroveň signálu v závislosti na vzdialenosti linky v oboch prípadoch [6]. Pre linku neovplyvnenou absorpciou O_2 je prenosová vzdialenosť 32 kilometrov. Inými slovami tento signál interferuje s akýmkoľvek iným signálom

rovnakej frekvencie na vzdialenosti 30 km za prijímačom. Počet liniek rovnakej frekvencie je prudko redukovaný pre spoločnú geografickú oblasť. Tiež to znamená, že nižšia úroveň signálu môže byť zachytená na vzdialenosti 30 km za prijímačom. V porovnaní s vysielacím signálom s frekvenciou 60 GHz, prenosová vzdialenosť linky je iba 2,5 km. Preto môže byť použitý oveľa väčší počet liniek v rovnakej geografickej oblasti bez toho aby boli rušené. 60 GHz linky sú preto oveľa bezpečnejšie z pohľadu na ich obmedzený rozsah.



Obr. 4 Útlm spôsobený kyslíkom v mm pásme vĺn.



Obr. 5 Vplyv absorpcie O_2 na šíriaci sa signál.

Smerovosť antény

Smerovosť je meradlom toho, ako dobre dokáže anténa sústrediť svoju energiu na určený smer. Point-to-point spojenie by malo mať vysokú smerovosť antén, pretože cieľom je prepojiť koncové body linky. V ideálnom prípade všetka vysielaná energia je smerujúca len na určitého príjemcu. Vysoko sústredené antény minimalizujú množstvo rušenia medzi linkami v rovnakej geografickej oblasti, minimalizujú riziko, že prenos bude zachytený a maximalizujú výkon. Smerovosť antény je obmedzená fyzikálnym princípom difrakcie, ktorá hovorí, že šírka lúča je priamo úmerná pracovnej frekvencii. Preto pri frekvencii 60 GHz šírka zväzku lúča je ďaleko užšia než u nižších frekvenčných pásmach. V *Tab.2* [6] je zobrazená šírka lúča pre niekoľko nelicencovaných frekvenčných pásiem. Tieto výsledky sú graficky znázornené na *Obr. 6* v ďalšej časti.

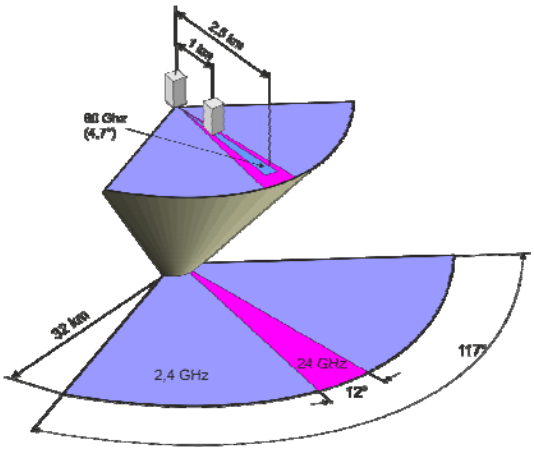
Tabuľka 2
ŠÍRKA LÚČA PRE NELICENCOVANÉ FREKVENČNÉ PÁSMA S PRIEMEROM ANTÉN 30,48 CM

Frekvencia	Šírka lúča
2,4 GHz	117 stupňov
24 GHz	12 stupňov
60 GHz	4,7 stupňov

Limitovanie žiarenia pre 60 GHz frekvencie

Kombináciou absorpcie O_2 a úzkeho zväzku lúča získame vysokú bezpečnosť prenosovej linky, viacnásobné využívanie frekvencie a nízke rušenie tejto 60 GHz linky. *Obr. 6* zobrazuje dve budovy, ktoré sú vzdialené od seba 1 km. Kliny zobrazujú vyžarovací diagram pre 2,4; 24 a 60 GHz. Linky pracujú s rovnakou prenosovou vzdialenosťou a to 1 km a využívajú antény

s priemerom 30,48 cm. Tri kliny zobrazujú miesta, kde žiarenie každej frekvencie je najväčšie. Najväčší klin predstavuje vyžarovací diagram z linky 2,4 GHz. 60 GHz linka má najužší klin. Praktické dôsledky sú zrejmé z *Obr. 6* [6].



Obr. 6 Diagram šírenia pre 2,4; 24 a 60 GHz linky.

60 GHz linka je interpretovaná úzkym klinom a nebude interferentovať s inými 60 GHz linkami. 24 GHz linka má rušenie a riziko odchytenia väčšie, taktiež jej klin je trochu širší. 2,4 GHz linka má veľké rušenie a riziko odchytenia je vo veľmi širokom okolí a to ako do vzdialenosti tak aj do šírky.

IV. ZÁVER

Hybridná FSO/RF linka využíva silné vlastnosti oboch liniek na prekonanie ich slabých stránok. Tým môže táto hybridná FSO/RF linka dosiahnuť vyššiu dostupnosť prenosovej linky ako je dosiahnutá samostatnou FSO linkou alebo RF linkou a to kvôli tomu, že RF časť hybridného FSO/RF môže zmierniť vplyv hustej hmly a FSO časť môže zmierniť vplyv dažďových zrážok. Hmla je najdôležitejším nepriaznivým faktorom FSO spojenia a dážď je najdôležitejším nepriaznivým faktorom RF spojenia. Výhody FSO linky pri jej zlyhaní nemusia byť stratené a to využitím 60 GHz záložnej linky, ktorá svojimi výhodami sa najviac približuje FSO linke

POĎAKOVANIE

Podporujeme výskumné aktivity na Slovensku/ Projekt je spolufinancovaný zo zdrojov EÚ. Tento článok bol vypracovaný v rámci projektu "Centrum excelentnosti integrovaného výskumu a využitia progresívnych materiálov a technológií v oblasti automobilovej elektroniky", ITMS 26220120055 (50%). Táto práca bola taktiež podporovaná a realizovaná z projektu Rozvoj Centra informačných a komunikačných technológií pre znalostné systémy (ITMS kód: 26220120030) na základe podpory operačného programu Výskum a vývoj financovaného z Európskeho fondu regionálneho rozvoja (50%).

LITERATÚRA

[1] L. C. Andrews, C. Y. Philips, J. Hopen, „Laser Beam Scintilation with Applications“, *SPIE Press.*, Washington , Bellingham, 2001.

[2] K. Chris, „The Benefits of 60 GHz Unlicensed Wireless Communications“. Dostupné na internete: http://www.ydi.com/products/fiberleap.php

[3] E. Leitgeb, & al., „Weather effects on hybrid FSO/RF communication link,…“, *IEEE J.Select. Areas in Comm.*, vol. 27, no. 9, December 2009, pp. 1687-1697.

[4] R. Luna, D. K. Borah, H. Tapse, „Behavior of hybrid optical/RF channels...“, *Proc. IEEE GLOBECOM*, Honolulu, HA, November-December 2009.

[5] H. Willebrand, B.S. Ghuman, „Free Space Optics: Enabling Optical Connectivity in Today´s Network“, *Sams Publishing*, 2002.

[6] J. Winters, „On the capacity of radio communication systems with diversity in Rayleigh fading enviroment“, *IEEE J.Select. Areas Communications*, vol. 5, June 1987, pp. 871-878.

Prepínanie RF a FSO prenosovej linky hybridného FSO/RF systému

Pavol MIŠENČÍK, Ľuboš OVSENÍK, Ján TURÁN

Katedra Elektroniky a multimediálnych telekomunikácií, Fakulta Elektrotechniky a Informatiky, Technická Univerzita v Košiciach, Slovenská Republika

pavol.misencik@tuke.sk, lubos.ovsenik@tuke.sk, jan.turan@tuke.sk

***Abstrakt* — Tento článok sa zaoberá prepínaním v hybridnom FSO/RF systéme. V prvej časti je popísaný hybridný FSO/RF systém a druhy prepínania medzi FSO a RF linkou. V ďalšej publikácie je rozpracovaný samotný návrh prepínania medzi FSO a RF linkou, ktorý je realizovaný pomocou prepínača s využitím RSTP protokolu. V závere je vypočítaná dostupnosť spojená s daným prepínaním.**

***Kľúčové slová* — FSO systém, hybridný FSO/RF systém, vlastnosti hybridného FSO/RF systému, MIMO prenosový systém**

I. Úvod

Komunikácia pomocou širokopásmovej technológie FSO sa začleňuje do klasickej komunikačnej siete len veľmi pomaly [1]. Hoci jej výhody oproti klasickej vláknovej optickej sieti sú značné. Tento fakt je spôsobený nepriaznivými vplyvmi atmosférického rušenia na spoľahlivosť FSO prenosu. Keďže vplyvy závisia od počasia, sú náhodné a dlhodobo nepredvídateľné. Náhodný vplyv na optický lúč môže byť od 0,2 dB pri úplne jasnom počasí, až po 350 dB pri hustej hmle. To spôsobuje, že spojenie potom nedosahuje požadovanú dostupnosť 99,999 % [1],[4].

Našou snahou je tento negatívny nedostatok minimalizovať alebo vyriešiť použitím záložnej prenosovej cesty. Pričom toto riešenie musí mať lepšiu dostupnosť ako klasický FSO prenos aj pri extrémnych poveternostných podmienkach [1],[4].

Najlepším riešením na zvýšenie dostupnosti, ale taktiež riešenie problému v prípade úplného výpadku FSO spojenia, je hybridný FSO/RF systém. Ako už z názvu vyplýva, tento hybridný systém obsahuje okrem primárneho optického prenosu aj záložný rádio–frekvenčný prenos [1].

II. Charakteristika FSO a RF Linky

A. FSO prenosová linka

Najčastejšie využívaná modulácia pre optické prenosy je modulácia s priamou detekciou. Z toho dôvodu FSO používa On – Off Keying (OOK) moduláciu, pričom takto modulované dáta sú priamo privádzané na vysielací laser. Na prijímacej strane fotodióda prevádza prijatý optický signál na elektrický a ten je potom spätne demodulovaný a ďalej spracovaný. Prijatý signál y_1 vieme vyjadriť ako:

$$y_1 = h\gamma x_0 + \eta_0 \tag{1}$$

kde y_1 je prijatý signál, h je kanálový zisk, γ označuje zisk fotoelektrického meniča (v jednotkách Ampér/Watt), x_0 reprezentuje intenzitu vysielaného optického signálu a η_0 je šum [1].

Najväčší negatívny vplyv na FSO kanál má hmla. Veľmi hustá hmla, pri ktorej klesá viditeľnosť pod 0,5 km, spôsobuje veľmi veľký útlm, čo má za následok aj prerušenie samotného spojenia, no závisí to aj od požadovanej dĺžky spojenia a vlnovej dĺžky optického signálu [2].

B. RF prenosová linka

Rádio – frekvenčný prenos ako záložný prenos v prípade výpadku FSO spojenia môže

pracovať vo viacerých frekvenčných pásmach, ktoré vyhovujú širokopásmovému prenosu. Keďže FSO pracuje v nelicencovanom pásme, tak sa uprednostňujú rádiové frekvencie, ktoré sú tiež v nelicencovanom pásme. Z tohto dôvodu sa na prenos používajú frekvencie 2,4 GHz, 5 GHz, 60 GHz, 80 GHz. V posledných dvoch pásmach sa zatiaľ dajú robiť prenosy len do jedného kilometra, čo môže byť aj výhoda [2]. Keďže sa jedná o milimetrové vlny, vplyv hmly nemá také deštrukčné účinky na prenosový kanál ako pri optickom kanáli. Útlm spôsobuje dážď, pretože veľkosť kvapôčok dažďa je porovnateľná s vlnovou dĺžkou RF signálu. Naproti tomu dážď vplýva zanedbateľne na optický signál. Z toho vyplýva, že RF a FSO prenosové cesty sa navzájom „dopĺňajú“, keďže je veľmi malá pravdepodobnosť výskytu hustej hmly a silného dažďa súčasne.

Modulácie, ktoré sa využívajú pri RF prenosovej linke sú: On–Off Keying (OOK), Binary Frequency–Shift Keying (BFSK) a Quadrature Phase–Shift Keying (QPSK) [2,6].

Prijatý výkon sa dá vypočítať pomocou Friisovho vzťahu:

$$P_r = \left(\frac{\lambda}{4\pi d}\right)^2 \frac{P_t G_t G_r}{L} \tag{2}$$

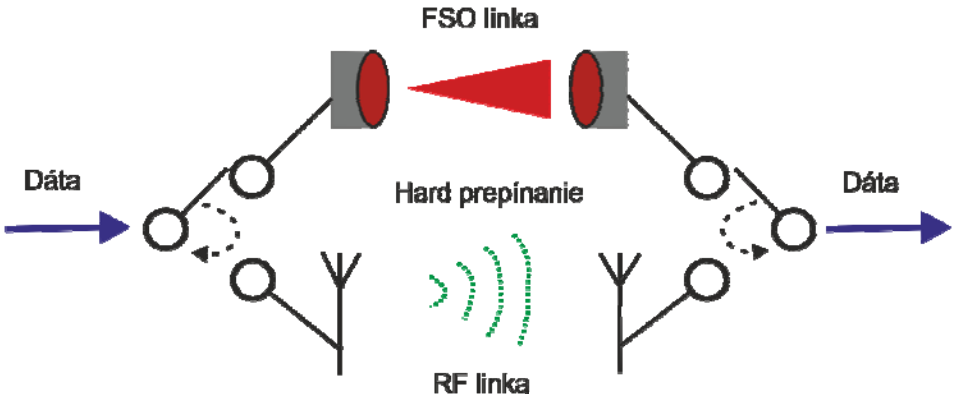
kde P_r je výkon prijatého signálu, λ je vlnová dĺžka signálu, d je vzdialenosť medzi vysielacou a prijímacou anténou, P_t je výkon vysielaného signálu, G_t a G_r reprezentujú zisk vysielacej a prijímacej antény a L sú systémové straty [6].

III. Prepínanie v Hybridnom Systéme

Ako už bolo spomenuté, hybridný systém obsahuje dve prenosové linky. Jednu primárnu, ktorú tvorí FSO linka a druhú záložnú RF linku. Preto je dôležite určiť ako systém prepína medzi týmito linkami. Systém používa jednu z dvoch metód prepínania.

A. Tvrdé prepínanie (Hard switching)

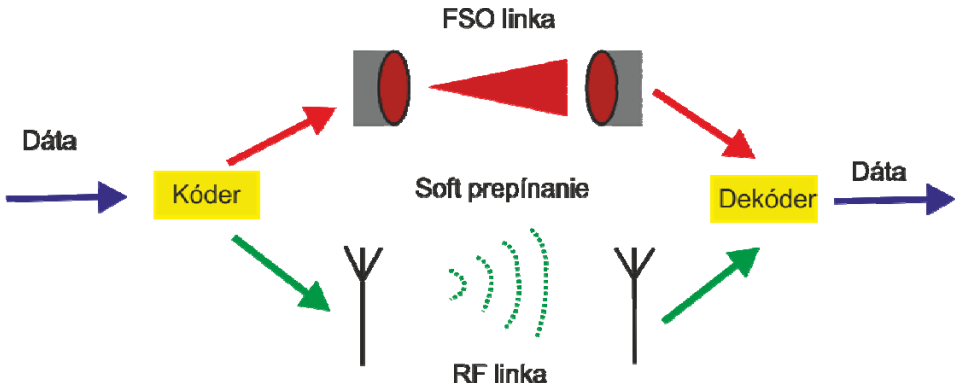
Pri tomto type prepínania musia vysielač a prijímač synchrónne vyberať medzi FSO a RF prenosovou linkou, pričom v určitom čase vysiela len jedna, tak ako je to zobrazené na *Obr. 1*. FSO linka je vybraná primárne a iba v prípade výpadku FSO linky sa prepne na RF linku. Spätne prepnutie nastane, ak sa obnoví spoľahlivý FSO prenos. Pri tvrdom prepínaní musí medzi vysielačom a prijímačom bežať spätná väzba, ktorá skoordinuje prepínanie na oboch stranách. Na prepínanie sa používajú rôzne protokoly pracujúce na druhej a tretej vrstve OSI modelu [3].



Obr. 1 Tvrdé prepínanie (Hard switching).

B. Mäkké prepínanie (Soft switching)

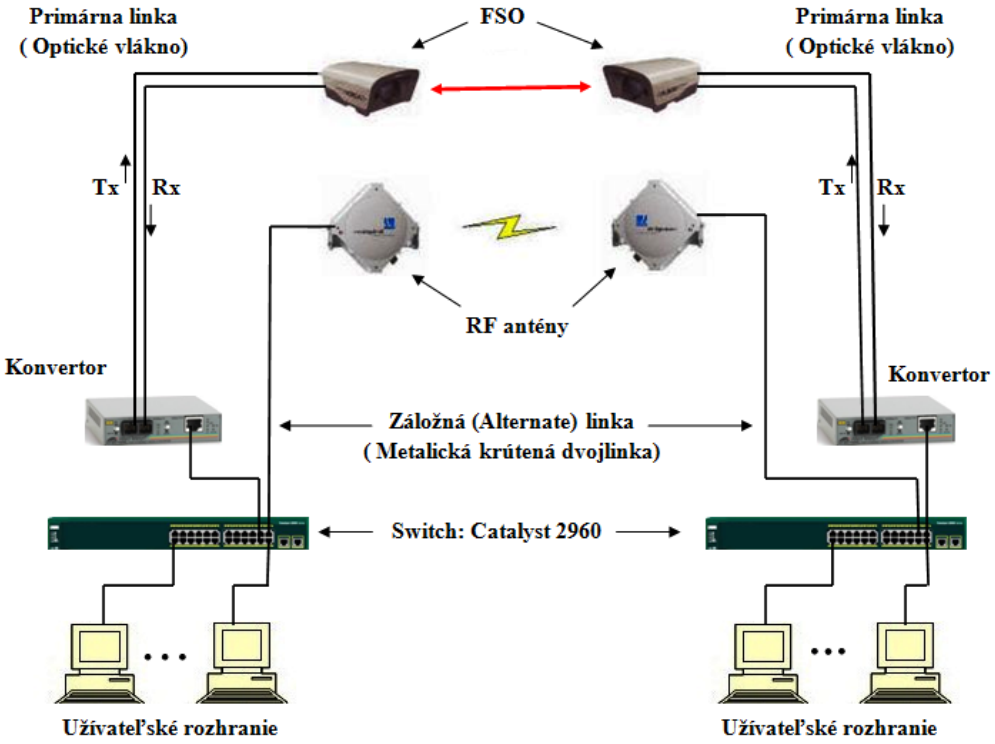
Pri tejto metóde sa namiesto prepínania medzi jednotlivými linkami používa kódovanie (napríklad Short-length Raptor kód) ako je to zobrazené na *Obr. 2*. Po zakódovaní sú dáta posielané súčasne oboma prenosovými cestami. FSO linka vysiela väčšou prenosovou kapacitou a RF linka menšou. Kódovanie však prispôsobuje prenos podľa podmienok vplývajúcich z prenosovej cesty.



Obr. 2 Mäkké prepínanie (Soft switching).

IV. NÁVRH PREPÍNAČA

Dôležitou časťou hybridnej FSO/RF linky je prepínač, ktorý vykonáva prepínanie medzi primárnou a záložnou linkou. Budeme a zaoberať tvrdým prepínaním medzi dvoma linkami. Primárnu linku tvorí FSO linka a záložnú RF linka. Prepínač musí zabezpečiť pri výpadku hlavnej FSO linky prepnutie komunikácie na záložnú RF linku a pri opätovnom obnovení FSO linky túto komunikáciu presmerovať späť na ňu. Dôležitou požiadavkou na prepínanie je čo najkratšia doba trvania, kým sa zistí výpadok primárnej linky a následne sa prepne na záložnú linku. Tento časový prírastok sa premietne v celkovej dostupnosti celej hybridnej FSO/RF linke. Na *Obr. 3* je znázornený konkrétny návrh hybridnej linky.



Obr. 3 Hybridný FSO/RF systém s prepínaním.

Jednotlivé časti hybridného systému tvoria FSO hlavice, ktoré sú pomocou dvoch optických vlákien pripojené na konvertor. Dvojitosť optických vlákien je kvôli plne duplexnému prenosu. Konvertor slúži na premenu optického signálu na elektrický a opačne. Určité typy prepínačov majú tento konvertor zabudovaný už v sebe, v tom prípade nie je nutné používať externý konvertor. Pri tomto návrhu sa však používa Cisco switch Catalyst 2960, ktorý obsahuje 24 Ethernet portov, 2 uplink porty a je bez optického rozhrania. Konvertor je pripojený pomocou metalického kábla do prepínača. Z prepínača okrem primárnej linky vedie aj záložná linka, ktorá je pripojená do RF antény. Antény používané na záložnú RF linku pracujú v nelicencovanom pásme. Tieto pásma môžu byť napríklad 2,4 GHz, 5,6 GHz alebo 60 GHz. Veľmi dôležité pri zapájaní primárnej a záložnej linky je dodržať pravidlo, že primárna linka sa

zapája do portu s nižším identifikátorom (nižšie číslo portu) ako záložná linka. Pravidlo vyplýva z použitého protokolu, ktorý je vysvetlený v nasledujúcej časti. Poslednou časťou tohto systému je užívateľské rozhranie, ktoré môže byť tvorené samotnými užívateľskými stanicami alebo ďalšou pripojenou sieťou.

RSTP Protokol (Rapid Spanning Tree Protocol)

Hybridný FSO/RF systém je redundantný systém s dvomi cestami do jedného cieľa. Použitý protokol v prepínači musí zamedziť súčasnej prevádzke obidvoch liniek a ak je to možné, tak zabezpečiť komunikáciu cez primárnu FSO linku. RSTP protokol je vylepšená a doplnená verzia STP protokolu. Veľkou nevýhodou STP protokolu je malá konvergencia, ktorá je veľmi dôležitá. Konvergentný čas STP protokolu môže byť od 30 až do 55 sekúnd pri prepnutí z blokovacieho stavu do stavu prenášaného. RSTP je typicky schopný zmenšiť tento čas na menej než 6 sekúnd, ak primárna linka zlyhá.

Hlavnou úlohou RSTP protokolu je aj predchádzať slučke. Protokol podrobne sleduje sieť a vyhľadáva všetky linky. Pritom vypnutím redundantnej linky zabezpečí, aby nevznikla žiadna slučka. Pomocou algoritmu kostry grafu (*STA - Spanning-Tree Algorithm*) protokol najprv vytvorí typologickú databázu a potom vyhľadá a zaistí redundantné linky. Ak je RSTP protokol aktívny, tak rámce prechádzajú iba cez najlepšie spoje, ktoré tento protokol vyberá.

RSTP pracuje, v porovnaní s STP, s pridaním alternatívneho portu a záložného portu. Tieto porty umožňujú ihneď prejsť na preposielajúci stav portu bez toho, aby pasívne čakali na konvergenciu siete.

Na komunikáciu medzi sebou si prepínače vymieňajú BPDU (*Bridge Protocol Data Unit*) správy. BPDU obsahujú identifikátor protokolu, typ BPDU, verziu protokolu, identifikátor root prepínača a vysielajúceho portu, identifikátor vysielajúceho prepínača, maximálnu dobu, oneskorenie pri posielaní a Hello time.

RSTP má k dispozícii iba tri stavy portov, ktoré zodpovedajú určitej činnosti:

Discarding (odmietajúci) – port neposiela rámce ďalej na sieť, prijíma BPDU rámce, ale neučí sa fyzickú adresu z BPDU rámcov ako pri STP blocking statuse.

Learning (učiaci sa) – prijíma a preposiela BPDU správy, učí sa fyzické adresy ale neposúva a neprijíma ešte užívateľské správy.

Forwarding (preposielajúci) – v tomto stave port prijíma a posiela užívateľské správy, normálna operácia portu, učí sa fyzické adresy, prijíma a posiela BPDU správy.

Dôležitými okrem stavov definovanými RSTP protokolmi sú aj úlohy portov. Najprv si musíme objasniť čo je identifikátor prepínača. Identifikátor je tvorený z fyzickej adresy (MAC) prepínača a taktiež z priority prepínača. Pomocou tohto identifikátora prepínače si určia Root prepínač a to tak, že Root prepínač sa stane prepínač s nižším identifikátorom. Na začiatku si každý prepínač myslí o sebe že je root. Po štarte si medzi sebou začnú vymieňať BPDU. Tieto BPDU majú v sebe obsiahnuté identifikátor root prepínača a identifikátor vysielacieho prepínača. Oba prepínače príjmu tieto BPDU, porovnajú svoje identifikátory s prijatými a určia, ktorý z nich je root prepínač.

Funkcie portov, ktoré môžeme rozlišovať pri RSTP protokole:

Root port - je port prepínača, z ktorého vedie najlepšia cesta (najlepšia cesta je cesta s nižším súčtom path cost všetkých liniek (*Tab. 1*). V našom prípade to bude FSO cesta.) k Root prepínaču. V prípade, že na prepínači existujú rovnaké porty, odkiaľ vedie najkratšia cesta, vyberie sa port, ktorý prijíma správy od prepínača s nižším identifikátorom. Ak aj pri tomto porovnaní existuje viac zhodných portov, vyberie sa port, ktorý prijíma správy od portu s lepším indentifikátorom.

Designated port - je port na linke, z ktorého vedie najlepšia cesta k Root prepínaču. Je to port presne na opačnej strane linky, kde je Root port. V prípade, že existujú porty s rovnakými hodnotami vyberie sa port podľa postupu ako pri Root porte.

Alternate port - je to blokovaný port, avšak v prípade stratenia spojenia k Root prepínaču môže sa stať aj Root portom. Je to port, ktorý môže vytvoriť alternatívnu cestu k Root prepínaču.

Backup port - je tiež blokovaný port. Tento port prijíma BPDU z toho istého prepínača ako je on sám. Je to port, ktorý vytvára záložnú cestu.

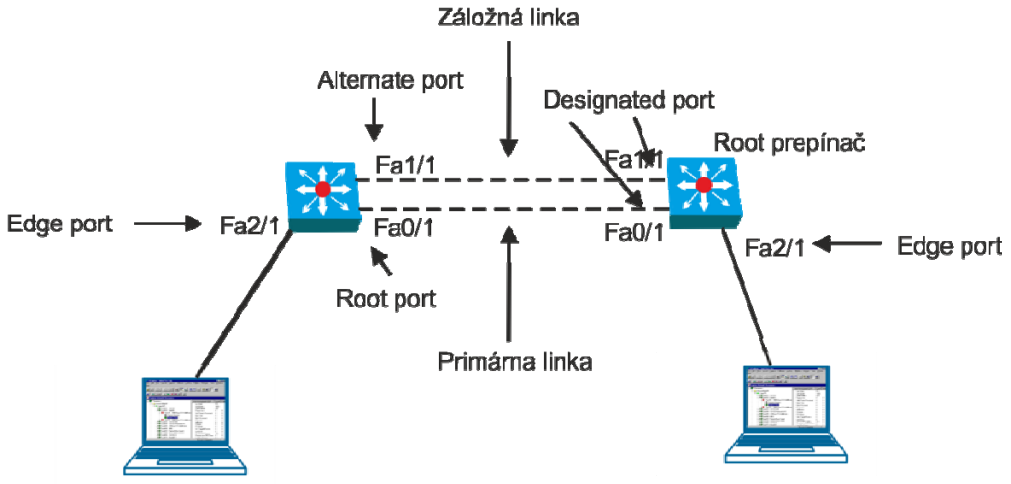
RSTP umožňuje rýchle obnovenie pripojenia po zlyhaní portu prepínača alebo LAN. To umožňuje rýchlu konvergenciu pre nové root porty a porty pripojené cez point-to-point pripojenie. Ak RSTP zvolí nový root port, blokuje starý root port a okamžite nový root port prechádza do „forwarding“ stavu. Teda v našom zapojení máme iba dve linky, ktoré sú pripojené na dva porty. Do root portu je pripojená FSO linka, ktorá je primárnou linku vďaka jej lepšej „path cost“ čiže šírke pásma (*Tab. 1*). Prepínač stále prijíma BPDU správy po oboch linkách. Ak prijímač neprijme trikrát PBDU na root port, tak to vyhodnotí ako stratu spojenia. Celú komunikáciu okamžite prepne na svoj „alternate“ port (*Obr. 4*), ktorý automaticky mení

svoj „discarding“ stav na „forwarding“. Hneď potom prebieha synchronizácia, po ktorej dochádza k zmene „alternate“ portu na „root“ port.

Po obnovení primárnej linky prepínače si začnú vymieňať BPDU po oboch linkách, pretože porty na primárnej linke zostali stále v stave „forwarding“. Obidva porty nachádzajúce sa na primárnej linke sa nastavia do stavu „discarding“. „Root prepínač pošle po novej (primárnej) linke „proposal“ BPDU (návrh) na suseda. Druhý prepínač zistí, že obnovená linka obsahuje „kratšiu“ cestu k root prepínaču, a preto svoje porty, okrem „edge“ portov, zablokuje. Hneď vyšle „agreement BPDU“ správu, čím sa hneď obnoví root port a tým aj primárna FSO linka. Podobná komunikácia prebehne aj na druhej linke. Zistí sa „dlhšia“ cesta a zmení sa port na „alternate port“, ktorý prejde do stavu „discarding“, čím vzniká opäť záložná linka. Čas potrebný na takúto konvergenciu je 6 sekúnd a vyplýva to z 2 sekundových intervaloch posielania PBDU. Tento čas zohľadníme v závere pri výpočte dostupnosti paralelného systému.

Tabuľka 1
Hodnota Path cost v závislosti od kapacity linky

Rýchlosť linky	Odporúčaná hodnota	Odporúčaný rozsah hodnôt
100 K	200 000 0 00	20 000 000 – 200 000 000
1 Mb	20 000 000	2 000 000 – 200 000 000
10 Mb	2 000 000	200 000 – 20 000 000
100 Mb	200 000	20 000 – 2000 000
1 Gb	20 000	2 000 – 200 000
10 Gb	2 000	200 – 20 000
100 Gb	200	20 – 2 000
1 Tb	20	2 – 200
10 Tb	0	1 - 20



Obr. 4 Úlohy portov v RSTP protokole.

V. DOSTUPNOSŤ SYSTÉMU

Na celkovú dostupnosť linky vplýva do značnej miery okolité prostredie, kde sa daný prenosový systém nachádza a taktiež vzdialenosť, na ktorú linka operuje. Z uvedeného dôvodu sa vypočítaná dostupnosť nedá použiť univerzálne pre každý systém nachádzajúci sa na rôznych miestach. Preto pri počítaní dostupnosti hybridného FSO/RF systému je potrebné určiť niekoľko podmienok:

1. Aký systém na prenos využijeme (dôležité sú jeho technické parametre).
2. Akú záložnú RF linku použijeme (s akou dostupnosťou v danej oblasti).
3. Stanoviť vzdialenosť prenosovej hybridnej linky.

Hodnota dostupnosti vyjadruje mieru funkčnosti systému. Na výpočet dostupnosti D použijeme vzťah:

$$D = \left(1 - \frac{\textit{čas výpadkov}}{\textit{celkový čas merania}}\right) \cdot 100\% \qquad (3)$$

Za celkový čas merania sa vo väčšine prípadov považuje obdobie jedného roka. Po prepočítaní na hodiny (365 x 24) to vychádza na 8760 hodín. Nasledujúca tabuľka zobrazuje niektoré hodnoty dostupnosti a k ním patriace časy prevádzky a výpadkov [4].

Tabuľka 2
Hodnota Path cost v závislosti od kapacity linky

Dostupnosť v %	Čas prevádzky [h]	Čas výpadkov
50	4380	4380 hodín
80	7008	1752 hodín
90	7884	876 hodín
95	8322	438 hodín
98	8585	175 hodín
99	8672	88 hodín
99,5	8716,2	43,8 hodín
99,9	8751,24	8,8 hodín
99,99	8759,124	53 minút
99,999	8759,912	5,3 minút
99,9999	8759,991	32 sekúnd

Podrobné určenie dostupnosti v Košiciach je určené v [4].

VI. ZÁVER

Uviedli sme základné riešenia problémov spojených s prepínačom, ktorý v hybridnom FSO/RF systéme zabezpečuje dostupnosť celkovej komunikačnej linky. Jeho úlohou je zabezpečiť pri zlyhaní primárnej linky(FSO) presmerovanie celkovej komunikácie na záložnú linku (RF). Celkové oneskorenie, ktoré tento prepínač vnesie do hybridného prenosového systému je 6 sekúnd, čo je 0,00166 hodín, pri každom prepnutí. V práci [4] je určená dostupnosť FSO systému vďaka meraným údajom viditeľnosti. Hodnota tejto dostupnosti je 98,4132 %. Podľa nameraných hodnôt 139 hodín do roka je FSO linka nefunkčná a dôjde k prepnutiu na záložnú RF linku 37 krát. Vypočítaná dostupnosť hybridného FSO/RF systému je 98,4125 %, čo ukazuje, že rýchlosť detekcie výpadku 6 sekúnd je pre tento systém postačujúca.

POĎAKOVANIE

Podporujeme výskumné aktivity na Slovensku/ Projekt je spolufinancovaný zo zdrojov EÚ. Tento článok bol vypracovaný v rámci projektu "Centrum excelentnosti integrovaného výskumu a využitia progresívnych materiálov a technológií v oblasti automobilovej elektroniky", ITMS 26220120055 (50%). Táto práca bola taktiež podporovaná a realizovaná z projektu Rozvoj Centra informačných a komunikačných technológií pre znalostné systémy (ITMS kód: 26220120030) na základe podpory operačného programu Výskum a vývoj financovaného z Európskeho fondu regionálneho rozvoja (50%).

LITERATÚRA

[1] I. E. Lee, Z. Ghassemlooy, Ng, Wai-Pang, S. Rajbhandari, "Fundamental Analysis of Hybrid Free Space Optical and Radio Frequency Communication Systems",Northumbria University, United Kingdom.
[2] P. Mišenčík, T. Harasthy, J. Turán, "Availability of Hybrid FSO/RF System. Department of Electronics and Multimedia Communications", FEI TU of Košice, Slovak Republic.
[3] W. Zhang, S. Hranilovic, C. Shi, „Soft-Switching Hybrid FSO/RF Linka using Short-length Raptor codes: Design and Implementation", *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 9, December 2009.
[4] M. Redaj, *Prepínanie RF a FSO prenosovej linky hybridného FSO/RF systému* (Bakalárska práca). Technická univerzita v Košiciach, Fakulta elektrotechniky a informatiky, 2012.
[5] I. I. Kim, E. Korevaar, „Availability of Free Space Optics (FSO) and Hybrid FSO/RF Systems", *Proc. SPIE*, , vol. 4530, 2001, pp. 84–95.
[6] Š. Vávra, J. Turán, *Antény a šírenie elektromagnetických vĺn*. Alfa, Bratilava, 1990.

Slepá steganalýza v statických obrazoch

Daniel MAJERČÁK, Vladimír BÁNOCI, Gabriel BUGÁR, Dušan LEVICKÝ

Katedra elektroniky a multimediálnych telekomunikácií, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

daniel.majercak@student.tuke.sk, vladimir.banoci@tuke.sk, gabriel.bugar@tuke.sk, dusan.levicky@tuke.sk

***Abstrakt* — Predložený článok sa zaoberá problematikou steganalýzy t.j. detekciou utajenej komunikácie prostredníctvom voľne dostupných steganografických metód aplikovaných v statických obrazoch vo formáte JPEG. Obsahuje návrh modulu univerzálnej steganalýzy, využívajúceho transformovanú DCT oblasť, ktorý je schopný rozpoznať najviac používané steganografické algoritmy, vyskytujúce sa na internete, ako sú metódy Jsteg, F5, Outguess. Článok sa taktiež venuje problematike návrhu steganalýzneho systému, ako aj štatistickým charakteristikám v statických obrazoch, ktoré sú nevyhnutné pri modelovaní takéhoto systému.**

***Kľúčové slová* — steganalýza, histogram, DCT, klasifikácia, Kalibrácia**

I. ÚVOD

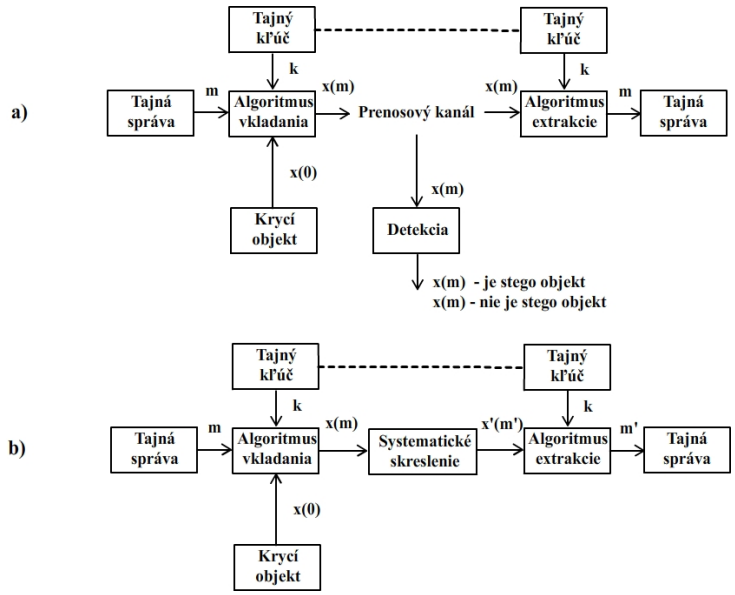
V posledných rokoch sme mohli zaznamenať výrazný nárast digitalizácie rôznych druhov informácií. Na druhej strane sa takáto forma reprezentácie stáva živnou pôdou pre nelegálnu činnosť alebo činnosť,ktorá ohrozuje spoločnosť. Ide napríklad o aktivity organizujúce teroristické útoky, plány krádeží a hackerské útoky s cieľom poškodenia informačných a komunikačných systémov.

Z tohto pohľadu je zaujímavá oblasť steganografie, ktorá realizuje utajenú komunikáciu na pozadí verejnej, čiže neutajenej komunikácie. So zvyšujúcimi sa obavami zo strany verejného a súkromného sektora, začali vznikať požiadavky na kontrolu tejto utajenej činnosti, ktorých výsledkom je steganalýza čo je veda zaoberajúca sa odhaľovaním podprahových kanálov v elektronických médiách a ich následným narušením. Napomáha monitorovaniu sieťových aktivít s možnosťou v predstihu zasiahnuť voči črtajúcej sa hrozbe.

Ak sa v procese steganalýzy podarí odhaliť utajenú komunikáciu, považuje sa takýto steganografický systém za prelomený a účel steganografie za zmarený [1]. Steganalýzna metóda sa považuje za úspešnú, ak dokáže rozlíšiť stego objekt od krycieho s väčšou pravdepodobnosťou ako je náhodné hádanie [2]. Steganografické vkladania a extrakcie sú výpočtovo jednoduché oproti spoľahlivej detekcii [3].

Steganalýza rozoznáva dva druhy útokov resp. dva druhy režimov [3]. Prvý druh útoku je pasívny. Osoba, ktorá vykonáva steganalýzu má právo iba čítať obsah prijatej správy a nie ho pozmeňovať. Jej cieľom je správne definovať existenciu podprahového kanála, prípadne určiť tajnú správu, pričom,poskytuje dôkaz, či daný x(i) je stego objekt alebo nie. Na základe neho je možné dokázať tretej strane, že použil steganografiu (Obr. 1a).

Druhým druh útoku je aktívny. Osoba vykonávajúca steganalýzu má právo vstupovať do komunikačného kanála. Jej cieľom je zabrániť možnej utajenej komunikácie pomocou narušiteľnej funkcie. Vykonáva systematické skreslenie v komunikačnom kanáli s cieľom narušiť alebo úplne zničiť potenciálnu tajnú správu. Skreslením možno útočiť na robustnosť steganografického systému. Nevýhodou je vedľajší efekt, ktorý sa prejaví ako šum pri skreslený objektu x(i), ktorý neobsahoval tajnú správu (Obr. 1b).



Obr. 1 a) Aktívny útok v steganalýze, b) Pasívny útok v steganalýze

II. OBRAZOVÁ STEGANALÝZA

Všetky prirodzené statické obrazy obsahujú veľa korelácie medzi susediacimi vzorkami, čím vykazujú určité štatistické vlastnosti [4]. Procesom vkladania tajnej správy sa tieto vlastnosti narúšajú a odchyľujú sa od určitej normy. Tento fakt je využívaný pri obrazovej steganalýze, ktorá sa snaží tieto zmeny zachytiť a na ich základe rozlíšiť stego obraz od krycieho obrazu. V posledných rokoch dochádzalo k rozvoju steganografických techník, ktoré boli verne nasledované vylepšenými steganalýznymi technikami pre ich detekciu. Ak nepovolaná osoba nevie nič o použitej steganografickej schéme, nazýva sa táto steganalýza ako univerzálna alebo slepá.

Univerzálne (slepé) steganalýzne metódy predstavujú súbor techník, ktoré sú nezávislé od použitého steganografického algoritmu, resp. dosahujú dobré výsledky detekcie tajnej správy, ktorá bola ukrytá novým alebo nepoznaným algoritmom.

Špeciálne steganalýzne metódy. Za predpokladu, že poznáme rozdelenie pravdepodobnosti krycích obrazov P_c, steganografický režim, a distribúciou správ, pomocou ktorých sa teoreticky vypočíta rozdelenie pravdepodobnosti stego obrazov P_s. Táto dodatočná informácia v podobe P_s dáva príležitosť útočiť na konkrétny algoritmus, a tým získať lepší detektor.

Základná myšlienka steganalýznych univerzálnych metód predpokladá, že prirodzený obraz možno definovať pomocou malej sady numerických funkcií [5]. Pre pokrytie celej distribúcie týchto funkcií je potrebná rozsiahla databáza snímok, ktorá je pomocou nej zmapovaná. V ďalšom kroku je následne nutné rozpoznať tieto vektory a zaradiť ich do príslušných tried resp. či skúmaný vektor patrí prirodzenému alebo stego obrazu. Prvýkrát prišiel s touto myšlienkou Avcibas et al. [1], ktorý testoval túto metódu na niekoľkých robustných algoritmoch digitálnej vodotlače, založených na vkladaní do LSB bitovej roviny, pričom meral metriku kvality obrazu (IQM)[5]. Neskôr navrhol sadu funkcií na základe binárnej podobnosti medzi siedmou a ôsmou bitovou rovinou. Ďalším autorom, ktorý navrhoval použiť sadu funkcií z vyššieho rádu štatistík obrazu pre klasifikáciu v steganalýze bol Farid [6]. Existuje veľa univerzálnych steganalýznych metód, medzi najzákladnejšie patria tie, ktoré sú založené na nižšie uvedených princípoch.

Binárne podobnostné funkcie – pôvodne publikované v Avcibs [1]. Sada funkcií je získaná z priestorovej reprezentácie obrazu. Autori tvrdia, že korelácia medzi susediacimi bitmi a bitmi jednotlivých rovín klesá po vložení tajnej správy. Konkrétne sa analyzuje siedma a ôsma bitová rovinu obrazu.

Steganalýza vo waveletovej oblasti – navrhnuté Lyu a Faridom [6]. Používajú štatistiky získané z obrazu vo waveletovej oblasti, pričom pri jeho rozklade využívajú QMF filtre. Z každej podoblasti a chyby predikcie, tzv. siedmich najdôležitejších susedov, sa vypočítavajú funkcie ako šikmosť, špicatosť, rozptyl a priemer.

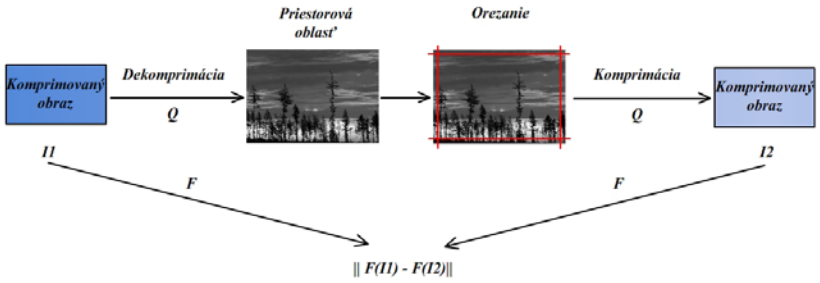
Kalibrované funkcie – tento súbor štatistík (Fridrich [7]) je získaný z DCT oblasti. Pričom sa snažia zachytiť rozdiel medzi pôvodným a kalibrovaným obrazom. Pod pojmom kalibrácia

rozumieme operáciu, ktorou sa snažíme priblížiť ku štatistikám blízkym pôvodnému prirodzenému obrazu.

A. Steganalýza založená na báze kalibrácie štatistík

Prvý univerzálny steganalýzny detektor, ktorý využíval kalibráciu 23 funkcií bol publikovaný J. Fridrich [7] a neskôr sa táto sada funkcií rozšírila na vektor o dĺžke 274. Kalibrácia bola pôvodne zavedená pre JPEG formát, ale je ju možné aplikovať aj v priestorovej oblasti [8]. Kalibrácia v JPEG obraze sa používa k odhadu štatistických vlastností krycieho obrazu zo stego obrazu. Tento proces je viac citlivý na zmeny spôsobené vkladaním tajnej správy ako na zmeny v obsahu obrazu. Jej nevýhoda spočíva vo väčšej citlivosti na faktor kvality stego objektu a ich kombináciou pri dvojito stlačených obrazoch. Tento problém je možné vyriešiť detekciou vybraných prípadov dvojitej kompresie pred vstupom do klasifikátora.

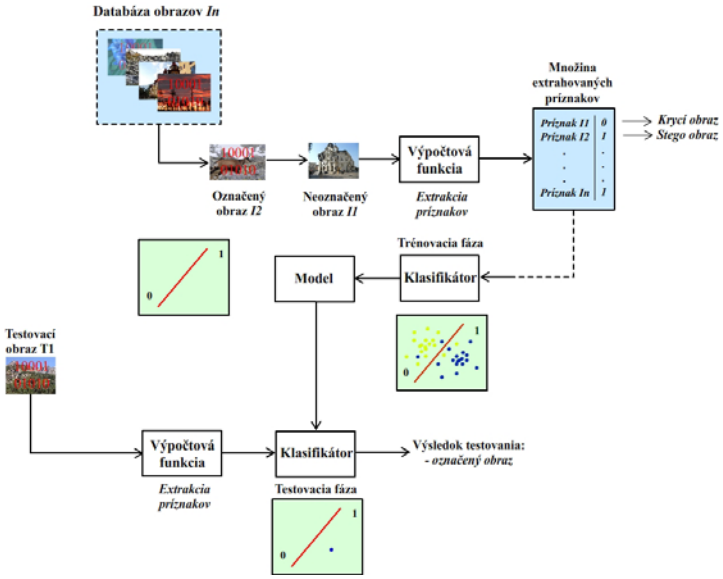
Pri kalibrácií komprimovaného obrazu dochádza najprv k dekomprimovaniu do priestorovej oblasti, kde je orezaný o štyri obrazové body v horizontálnom aj vertikálnom smere. Takto upravený obraz je následne komprimovaný pomocou rovnakej kvantizačnej matice do transformovanej oblasti (Obr. 2). Týmito operáciami získame obraz, ktorý je aj svojimi štatistickými vlastnosťami veľmi podobný originálnemu kryciemu obrazu.



Obr. 2 Proces kalibrácie v JPEG formáte spôsobený orezaním obrazu

III. NÁVRH MODULU UNIVERZÁLNEJ STEGANALÝZY

Modul univerzálnej steganalýzy vo všeobecnosti obsahuje veľkú databázu prirodzených obrazov a to označených aj neoznačených. Z nich sa podľa zvolenej metódy extrahujú príznaky, ktoré reprezentujú daný obraz. Na ich základe vie potom rozlíšiť, či skúmaný obraz je stego alebo krycí. Nasledujúci obrázok zobrazuje všeobecnú blokovú schému takéhoto modulu.



Obr. 3 Bloková schéma univerzálneho steganalyzéra

Z blokovej schémy univerzálneho steganalyzéra uvedeného na Obr. 3 vyplýva, že funkcionalita modulu sa skladá z troch fáz. Prvá fáza, ako už bolo spomínané, predstavuje

extrakciu príznakov z vytvorenej databázy. Tá je reprezentovaná jedným dôležitým blokom výpočtovej funkcie, ktorá predstavuje určitú univerzálnu steganalýznu metódu. Táto časť postupne spracováva, resp. mapuje všetky obrazy obsiahnuté v databáze a ukladá ich štatistické výsledky do zásobníka. V ňom je každému vektoru pridelená identifikácia, či sa pred spracovaním jednalo o stego obraz alebo neoznačený obraz, prípadne o akú steganografickú metódu išlo. Takto vytvorená sada funkcií následne postupuje do klasifikátora, ktorého úlohou je separovať príznaky patriace krycím obrazom od príznakov patriacich stego obrazom na základe prideleného identifikátora. Výsledkom operácie je natrénovaný model alebo modely, ktoré sú v tretej fáze schopné klasifikovať obrazy, prípadne ich začleniť do príslušných steganografických skupín. Natrenovanie dokonale univerzálneho modelu steganalýzy je náročná úloha. Niektorí autori tvrdia, že univerzálne steganalýzne metódy sú len zoskupením špeciálnych metód, ktoré potrebujú neustály proces vyvíjania sa, resp. natrénovania sa na novo vzniknuté algoritmy. Citlivosť, resp. presnosť takéhoto modelu ovplyvňuje veľký počet faktorov. Model by mal pokrývať celú distribúciu krycích obrazov, poprípade všetkých poznaných stego obrazov a taktiež závisí od druhu použitého klasifikátora a od voľby vstupných parametrov pri trénovaní modelu [9].

IV. Záver

Cieľom tohto príspevku bolo priblížiť problematiku steganalýzy, jej kategorizáciu z hľadiska detekcie často využívaných, či nových steganografických metód. Následne boli v článku popísané rôzne prístupy steganalýzy z hladiska detekcie a spracovania štatistík, ovplyvnených steganografikými metódami, zodpovednými za vkladanie tajnej informácie do statických obrazov. Pri návrhu uvedeného steganalytického modulu bola využitá oblasť DCT, a to v procese kalibrácie, ktorá sa používa k odhadu štatistických vlastností krycieho obrazu zo stego obrazu. Rozdielom medzi predikovaným krycím obrazom a pôvodným krycím obrazom dostávame dáta, na ktoré aplikujeme kalibrované štatistické funkcie a týmto dostávame charakteristický „odtlačok“, ktorý popisuje konkrétnu steganografickú metódu v procese vkladania tajnej správy. Jednou z najdôležitejších požiadaviek, ktorá je pri univerzálnej steganalýze veľmi žiaduca, je rozpoznanie nekonvenčných algoritmov. Týmto sa dostávame k problému výberu vhodných kalibračných štatistík, ako aj k správnemu natrénovaniu steganalýzneho modulu pre potreby detekcie nových, nepoznaných algoritmov.

Navrhnutý model univerzálnej steganalýzy, pracujúci v oblasti DCT transformácie, je systém veľmi flexibilný s možnosťou dodatočného natrénovania modelov a s detekciou rôznych druhov steganografických algoritmov. Nevýhoda metódy predstavuje jej zameranie na prirodzené obrazy a obrazy s menšími manipuláciami s obsahom, ktoré vznikli pri ich zachytení. Pri akejkoľvek väčšej digitálnej manipulácií s obrazom, klasifikuje navrhnutý modul takýto obraz za označený, prípadne klesá jeho výkon.

Poďakovanie

Táto práca bola vytvorená realizáciou projektu Rozvoj Centra informačných a komunikačných technológií pre znalostné systémy (kód ITMS projektu: 26220120030) na základe podpory operačného programu Výskum a vývoj financovaného z Európskeho fondu regionálneho rozvoja.

Literatúra

[1] AVCIBAS, Ismail – KHARRAZI, Mehdi – MEMON, Nasir – SANKUR, Búlent: Image steganalysis with binary similarity measures, Uludag University, Bursa, Turkey Detection in Steganalysis, Indian Institute of Science, Bangalore, june 2006

[2] CACHIN, C.: An information-theoretic model for steganography. In: Aucsmith, D. (ed.) Information Hiding (2nd International Workshop), LNCS 1525. Springer-Verlag, Berlin Heidelberg (1998).

[3] BOHMER Rainer: Advanced statistical Steganalysis. Dresden: Springer 2009. ISBN 978-3-642-14312-0

[4] MORKEL, T. – ELOFF, J.H.P. – OLIVIER, M.S. :An Overview of Image Steganography. Proceedings of the Fifth Annual Information Security South Africa Conference (ISSA2005)

[5] PEVNÝ Tomáš - FRIDRICH Jessica: Multi-class Blind Steganalysis for JPEG Images. Department of Computer Science, Binghamton, NY 13902-600.

[6] FARID, H. – SIWEI, L.:Detecting Hidden Messages Using Higher-Order Statistics and Support Vector Machines, in:F.A.P. Petitcolas (ed.): Information Hiding. 5th International Workshop. Lecture Notes in Computer Science, vol. 2578. Springer-Verlag, BerlinHeidelberg New York, 2002.

[7] FRIDRICH, Jessica – PEVNÝ Tomáš: Merging Markov and DCT Feature for Multi-Class JPEG Steganalysis. Department of Computer Science, Binghamton University, State University of New York.

[8] KER, A. D.: Steganalysis of LSB matching in grayscale images. IEEE Signal Processing Letters, 12(6):441–444, June 2005.

[9] LI, B. – HE, J. – HUANG, J. – SHI, Z. Q.: A Survey on Image Steganography and Steganalysis. Journal of Information Hiding and Multimedia Signal Processing, Volume 2, Number 2, April 2011, ISSN 2073-4212.

Distribuovaný dialógový manažér pre multimodálny dialógový systém

Stanislav ONDÁŠ, Martin VIŠŇOVSKÝ, Matej MOROZ

Katedra elektroniky a multimediálnych telekomunikácií, Fakulta elektrotechniky a informatiky, Technická Univerzita v Košiciach, Slovenská republika

stanislav.ondas@tuke.sk, martin.visnovsky@student.tuke.sk, matej.moroz@student.tuke.sk

***Abstrakt* — Predkladaný článok sa zaoberá návrhom multimodálneho dialógového systému a dialógového manažéra pre riadenie interakcie v takomto systéme. Navrhnutý multimodálny dialógový systém je založený na distribuovanej hub-server architektúre. Dialógový manažér nie je realizovaný ako jeden zo serverov systému, ale ako skupina agentov, ktorí spolu tvoria distribuovaný dialógový manažér. Špecifickým prvkom navrhnutého multimodálneho systému je dvojúrovňová komunikácia serverov, v ktorej sa hub architektúra využíva na vzájomné informovanie sa serverov (signalizáciu) a na zdieľanie dát súvisiacich s interakciou sa využíva spoločný dátový priestor vo forme databázy.**

***Kľúčové slová* — dialógový manažér, multimodálny dialógový systém, hub architektúra**

I. ÚVOD

Multimodálne dialógové systémy (MDS) sú interaktívne systémy, ktoré umožňujú používateľom interagovať s počítačovým systémom s využitím viacerých vstupno-výstupných modalít, ako sú reč, gestá, dotyk a pod. Multimodálna interakcia, ktorá zahŕňa vizuálne, zvukové, ako aj ďalšie zmyslové modality je najprirodzenejším typom komunikácie pre človeka. V súčasnosti sa stávajú dostupnými technológie, ktoré umožňujú aspoň limitované použitie niektorých modalít (gestá rúk a tváre, reč, mimika a pod.). Použitie viacerých modalít však prináša zvýšené nároky aj na ďalšie procesy v interaktívnych systémoch, ako sú fúzia a fízia modalít a v neposlednom rade aj na proces riadenia interakcie. Mnohé prístupy k riadeniu dialógu, ktoré sú dostatočne výkonné v relatívne jednoduchých úlohovo-orientovaných interakciách (napr. použitie dialógových jazykov, konečno-stavové automaty a pod.), sú príliš obmedzujúce pre umožnenie prirodzenejšej interakcie s používateľom.

Riadenie interakcie je teda stále živou problematikou, ktorej je potrebné sa venovať. Zároveň však môžeme konštatovať relatívne malý záujem o túto problematiku. Najväčšej popularite sa tešia tzv. jazyky na popis dialógu (Dialogue Description Languages – DDL), z ktorých najpoužívanejšími sú VoiceXML a SALT. Na interpretácii jazyka VoiceXML je založená aj dialógová jednotka navrhnutá a zrealizovaná v rámci vývoja systému IRKR (Inteligentné rečové komunikačné rozhranie) na našom pracovisku (pozri [4],[5]). Ako sme však uviedli vyššie, jazyk VoiceXML (verzie 1.0 až 2.1) sa javia ako limitujúce pre naozaj prirodzenú interakciu medzi používateľom a systémom. V tejto súvislosti sú veľké očakávania kladené na pripravovanú tretiu verziu jazyka VoiceXML (3.0) a riadiaceho jazyka SCXML, ktorá je však bohužiaľ už dlhšie obdobie iba v rozpracovanom stave.

V literatúre je možné nájsť viacero systémov so zaujímavými riešeniami riadenia dialógu, napr. [7], [3], [2], avšak často ide o riešenia, ktoré vznikli na akademickej pôde a sú buď nedokončené, chýba im potrebná dokumentácia, alebo nie sú voľne dostupné. Jedným z najzaujímavejších riešení predstavuje dialógová jednotka RavenClaw [7]. Ide o agentový prístup so zameraním na plánovanie. Žiaľ aj v tomto prípade chýba podrobnejšia dokumentácia. Navyše návrh aplikácie pre tento manažér je relatívne zložitý.

Na základe zváženia vyššie uvedených faktov a taktiež vychádzajúc zo skúseností, ktoré sme nadobudli pri návrhu a vývoji systému IRKR, sme sa rozhodli pre návrh vlastného riešenia MDS a jeho dialógového manažéra.

II. NÁVRH MULTIMODÁLNEHO DIALÓGOVÉHO SYSTÉMU

A. Východiská

Základnou požiadavkou, z ktorej vychádzajú všetky ostatné nároky na novo-budovaný systém bola čo možno najprirodzenejšia interakcia medzi človekom a systémom, teda čo najviac podobná komunikácii dvoch ľudí. Po analýze vlastností IRKR komunikátora a preštudovaní viacerých článkov ([7], [8], [9]) sme definovali tieto východiská a požiadavky na vlastnosti navrhovaného systému:

- Systém musí byť schopný kontinuálneho "vnímanie" používateľa, teda musí byť schopný kontinuálne spracovávať vstupný tok audio dát.
- Používateľ môže v každom okamžiku interakcie poskytnúť viac informácií, ako sa v danom kroku očakáva.
- Používateľ musí byť schopný kedykoľvek prerušiť výzvu systému a pokročiť v dialógu ďalej.
- Systém musí byť schopný nezávisle manažovať striedanie iniciatívy (tzv. turn-taking). To znamená, že po istom časovom úseku, keď užívateľ nereaguje sa slova ujíma systém a pokračuje sa v dialógu. Systém musí byť schopný rozpoznať, kedy mu používateľ predáva iniciatívu.
- Procesy systému musia byť asynchrónne s vykonávaním interakcie, ktorá musí byť synchrónna.

Na základe definovaných kritérií sme vylúčili prvotný návrh, v ktorom sme chceli na fyzickú realizáciu interakcie využiť VoiceXML interpreter, ktorý by vytváral finálnu podobu interakcie interpretovaním malých kúskov (poddialógov) hlasovej aplikácie. Jazyk VoiceXML vnímame ako prísne synchrónny, kde procesy riadenia interakcie a striedania iniciatívy sú striktne dané.

B. Architektúra a komponenty MDS

Na základe dlhoročných pozitívnych skúseností s distribuovanou hub-server architektúrou Galaxy [1], ktorá bola použitá pre vybudovanie rečového dialógového systému IRKR [4], ale aj jedného z najvýznamnejších akademických riešení rečového dialógového systému Olympus [6], sme sa rozhodli túto architektúru použiť ako základ navrhovaného multimodálneho systému.

Základná štruktúra navrhnutého MDS pozostáva z centrálneho riadiaceho prvku – tzv. Hub procesu a zo serverov, ktoré sú k nemu pripojené (viď Obr. 1). Servery systému je možne rozdeliť do dvoch skupín. Prvá skupina serverov (na Obr. 1. vľavo od zvislej červenej prerušovanej čiary) tvorí distribuovaný dialógový manažér. Druhá skupina serverov je zodpovedná za samotnú realizáciu interakcie a nazvali sme ju „Realizátory" (Realizers).

Servery distribuovaného dialógového manažéra sú:

- Manažér interakcie, Manažér dialógových zručností, Manažér úloh, Manažér používateľov, Manažér histórie, Informačný server,

Skupinu realizátorov interakcie tvoria:

- Vstupno/výstupný server (Input/Output interface), Automatický rozpoznávač reči (ASR server), Systém syntézy reči z textu (TTS server), Blok multimodálnej fúzie a analýzy modalít (Multimodal fusion and analysis), Blok extrahovania informácií z hlasu (Voice Info Engine),Manažér serverov

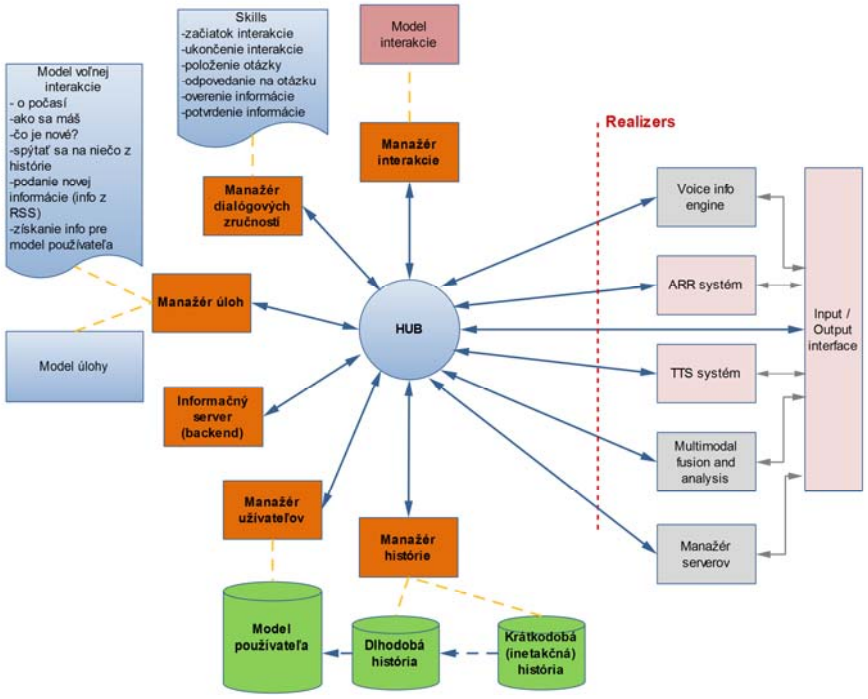
Z hľadiska typu architektúry je navrhnutý multimodálny dialógový systém systém založený na agentoch (manažéroch), s prístupom na báze plánovania, so spoločným dátovým priestorom vo forme databázy.

Základným mechanizmom riadenia interakcie je spúšťanie úloh na základe „spúšťačov" (triggerov) a dátových objektov rozpoznaných v reči používateľa. **Triggermi** označujeme skupina kľúčových slov rozpoznaných v odpovediach používateľa (zväčša slovesá a predmety), ktoré spúšťajú podúlohy interakcie. **Dátové objekty** sú zachytené kľúčové slová v rozpoznaných odpovediach používateľa, zväčša príslovkové určenia miesta, času, a pod., ktoré predstavujú informácie potrebné na zrealizovanie jednotlivých úloh interakcie.

Navrhnutá architektúra multimodálneho dialógového systému využíva dva spôsoby zdieľania informácií medzi komponentmi systému:

- **Správová komunikácia v Galaxy architektúra.** Slúži na informovanie sa o požiadavkách na služby, o priebehu služieb (realizovaných cez obslužné funkcie) a ich ukončení.
- **Spoločný priestor zdieľania dát** (Obr.2). Slúži na zdieľanie všetkých dát interakcie, teda informácií o úlohách, ich triggeroch, dátových objektoch, o používateľoch

a histórii interakcie, atď.

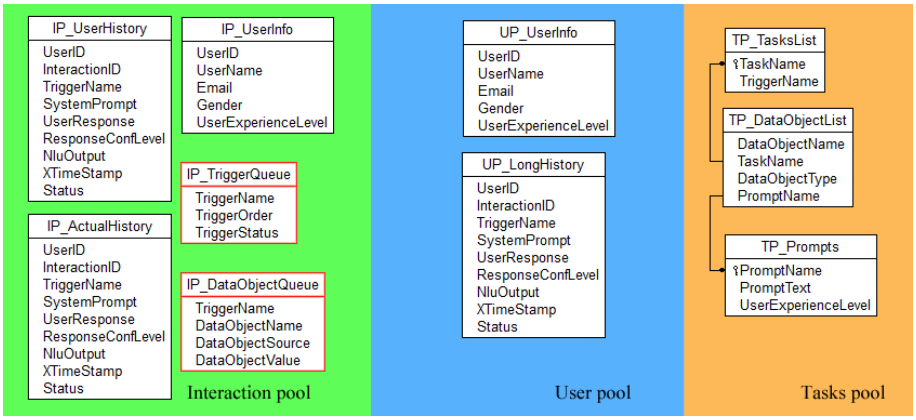


Obr. 1 Architektúra multimodálneho dialógového systému s distribuovaným dialógovým manažérom

Spoločný priestor zdieľania dát je tvorený databázou a je rozdelený na tri hlavné časti:

- Priestor s informáciami o interakcii (**Interaction pool**)
- Priestor s informáciami o používateľoch (**Users pool**)
- Priestor s informáciami o úlohách (**Tasks pool**)

Spoločný priestor zdieľania dát bude podrobnejšie popísaný v štvrtej kapitole.



Obr. 2 Priestor zdieľania dát v MDS

III. DISTRIBUOVANÝ DIALÓGOVÝ MANAŽÉR

Viacročný výskum a vývoj v oblasti rečových dialógových systémov a tiež navrhnutý a zrealizovaný dialógový manažér na báze interpretera jazyka VoiceXML [5] nám umožnili načerpať skúsenosti v tejto oblasti. Riešenie riadenia dialógov na báze jazyka VoiceXML je dobrým riešením pre jednoduché, dobre štruktúrované úlohy, avšak je ťažkopádnym až nemožným riešením pre voľnejšiu interakciu s atribútmi prirodzenej interakcie medzi ľuďmi. Preto sme sa rozhodli navrhnúť dialógový manažér, ktorý by umožňoval ovplyvňovať jednotlivé úlohy a procesy pri riadení interakcie medzi používateľom a systémom. Základnými kritériami pri návrhu bolo:

- zabezpečiť oddelenie dát, toku a prezentácie (data-flow-presentation), podobne, ako to navrhujú tvorcovia jazyka VoiceXML 3.0.
- zabezpečiť efektívnu a nenáročnú tvorbu nových aplikácií (úloh) pre takýto systém.

- zabezpečiť, aby procesy riadenia interakcie boli asynchrónne so samotnou interakciou.
- zabezpečiť možnosť personalizácie systému t.j. aby sa interakcia prispôsobovala konkrétnemu používateľovi a aby sa tieto nastavenia uložili a znovupoužili pri ďalšej interakcii s daným používateľom.

Na základe týchto kritérií sme v iteratívnom procese dospeli k riešeniu jednotky riadenia dialógov, ktoré je založené na kooperácii skupiny manažérov nad spoločným priestorom dát v podobe databázy. Ide teda o distribuovanú štruktúru serverov. Navrhnutý DM pozostáva z nasledujúcich manažérov:

- **Manažér interakcie** je hlavným manažérom DM. Podľa modelu interakcie riadi vykonávanie interakcie. Všetky informácie od ostatných manažérov, ako aj od *realizátorov* interakcie sú smerované na tohto manažéra.
- **Manažér úloh** je zodpovedný za prípravu a vykonávanie úloh, ktoré predstavujú služby alebo interakčné témy pre používateľa. Úloha je vnímaná ako súbor podúloh spúšťaných cez triggery a teda podúlohy sa nevykonávajú v pevne definovanom poradí. Špecifickým módom manažéra úloh je tzv. úloha voľnej interakcie. Tá môže slúžiť napr. na získanie informácií pre model používateľa.
- **Manažér dialógových zručností** je manažérom, ktorý operuje na prezentačnej vrstve systému. Je to zároveň jediný manažér, ktorý komunikuje s *realizátormi*. Ponúka škálu dialógových činností, ako privítanie, pýtanie sa, oznamovanie, potvrdzovanie, ukončenie interakcie a pod. Tieto dialógové zručnosti poskytuje manažérovi interakcie pre vytvorenie dialógu.
- **Manažér používateľov** spravuje informácie o používateľoch, ktoré umožňujú personalizovať systém a tak sa čo najviac prispôsobiť preferenciám používateľa.
- **Manažér histórie** spravuje krátkodobú (interakčnú) históriu a dlhodobú históriu daného používateľa. Je zodpovedný za transformáciu krátkodobej histórie do dlhodobej histórie používateľa.
- **Informačný server** získava informácii potrebných pre naplnenie úlohy systému a vykonáva ich spracovanie.

IV. Spoločný priestor zdieľania dát

Ako už bolo spomenuté, všetky informácie systému sú uložené v tzv. spoločnom priestore zdieľania dát, rozdelenom na tri časti - priestor s informáciami o interakcii (*Interaction pool*), o používateľoch (*Users pool*) a o úlohach (*Tasks pool*). Každý dátový priestor pozostáva z tabuliek, v ktorých sú informácie uložené. Tomuto riešeniu sme dali prednosť pred riešením založeným na reprezentácií informácií v samostatných súboroch, ktoré sú načítavané jednotlivými manažérmi a následne posielané cez hub. Dôvodom bola najmä obava zo zahltenia systému správami jednotlivých manažérov, ako aj z neprehľadnosti reprezentácie dát. Použitie databázových tabuliek je ľahko udržiavateľný spôsob reprezentácie dát. Ďalšou výhodou je fakt, že dáta takto uložené môžu byť jednoducho prehliadané aj systémovým správcom.

A. Priestor s informáciami o interakcii (Interaction pool)

Tento dátový priestor vzniká na začiatku interakcie s používateľom a pri jej skončení zaniká. Obsahuje dáta, ktoré súvisia s aktuálnou reláciou (interakciou) s používateľom. Pozostáva z nasledujúcich tabuliek :

- **IP_UserInfo** - obsahuje údaje o danom používateľovi. Na začiatku interakcie ju napĺňa *manažér používateľov*. Umožňuje personalizovať interakciu s používateľom.
- **IP_TriggersQueue** –zásobník (fronta) triggerov. Hlavný nástroj dialógového manažmentu. Interakčný manažér prehliada tento zásobník a na základe nájdených triggerov riadi interakciu.
- **IP_DataObjectQueue** – zásobník (fronta) dátových objektov. Každá úloha spúšťaná triggermi na seba viaže dátové objekty, ktoré je potrebné naplniť pre úspešnú realizáciu služby.
- **IP_ActualHistory** - obsahuje informácie o aktuálnej interakcii. Všetky akcie systému a používateľa sú zaznamenávané v tejto tabuľke a umožňujú dialógovému manažérovi pracovať s predošlými interakčnými udalosťami.
- **IP_UserHistory** - tabuľka obsahujúca informácie o predchádzajúcich interakciách daného používateľa so systémom. Umožňuje personalizáciu interakcie a použitie už skôr získaných informácií od používateľa.

B. *Priestor s informáciami o používateľoch (Users pool)*

Tento dátový priestor je trvalý a ostáva zachovaný aj po skončení interakcie. Uchováva informácie o používateľoch systému a histórii ich predošlých interakcií. Predstavuje pamäť celého systému.

- **UP_UserInfo** – tabuľka s informáciami o všetkých používateľoch.
- **UP_LongHistory** - tabuľka s dlhodobou históriou používateľov.

C. *Priestor s informáciami o úlohách (Tasks pool)*

Priestor s informáciami o úlohách uchováva informácie o úlohách, ich štruktúre, dátových objektoch a výzvach systému (prompts). Predstavuje úlohy, z ktorých pozostávajú služby poskytované systémom. Pozostáva z troch typov tabuliek:

- **TP_TasksList** – tabuľka jednotlivých úloh. Predstavuje zoznam úloh a ich triggerov.
- **TP_DataObjectList** – tabuľka jednotlivých dátových objektov. Predstavuje zoznam dátových objektov.
- **TP_Prompts** – tabuľka obsahujúca výzvy systému. Predstavuje zoznam výziev, ktoré môžu byť systémom prezentované.

V. ZÁVER

Predkladaný článok prezentuje prístup k návrhu moderného multimodálneho dialógového systému a dialógového manažéra pre riadenie interakcie v takomto systéme, ktorý sme použili pre vybudovanie multimodálneho dialógového systému, ktorý je zameraný na čo najprirodzenejšiu interakciu s používateľom. Hlavnou modalitou tohto systému je reč. Hlavnými atribútmi tohto návrhu je distribuovaný dialógový manažér, ktorý operuje nad spoločným priestorom dát realizovaným databázou s tromi priestormi zdieľania dát.

Prezentovaný návrh je toho času realizovaný iba čiastočne a preto naša ďalšia práca bude zameraná na dokončenie systému, vytvorenie pilotných aplikácií a na evaluáciu tohto systému.

POĎAKOVANIE

Táto práca bola vytvorená realizáciou projektu Rozvoj Centra informačných a komunikačných technológií pre znalostné systémy (kód ITMS projektu: 26220120030) na základe podpory operačného programu Výskum a vývoj financovaného z Európskeho fondu regionálneho rozvoja.

LITERATÚRA

[1] SENEFF, S. - HURLEY, E. - LAU, R. - PAO, C. - SCHMID, P. - ZUE, V.: Galaxy-II: a Reference Architecture for Conversational System Development. In Proceedings the 5th International Conference on Spoken Language Processing – of ICSLP'98, 1998.

[2] TURUNEN, M.: The Jaspis Architecture and Applications. Department of Computer and Informarion Sciences Tampere Unit for Computer-Human Interaction Speech-Based and Pervasive Interaction Group, COST 278 Management Committee meeting ETH, Zurich, Switzerland, February 4-5, 2002.

[3] WU, X., ZHENG, F., XU, M.: Topic Forest: A Plan-based Dialog Management Structure. In Proc. of the IEEE International Conference on Acoustics, Speech, and Signal Processing 2001 (ICASSP 2001), Salt Lake City, Utah.

[4] JUHÁR, J., a kol.: "Development of slovak GALAXY/voiceXML based spoken language dialogue system to retrieve information from the internet", In INTERSPEECH-2006, paper 2056-Mon2FoP.10.

[5] ONDÁŠ, S., JUHÁR, J.: Dialog manager based on the VoiceXML interpreter. In: Proceedings of the DSP-MCOM 2005 : The 6th international conference on Digital Signal Processing and Multimedia Communications : September 13-14, 2005, Košice, Slovac Republic. Košice : Technical university of Košice, 2005. s. 80-83. ISBN 80-8073-313-9.

[6] Bohus, Dan; Antoine Raux; Thomas K. Harris; Maxine Eskenazi & Alexander I. Rudnicky (2007), "Olympus: an open-source framework for conversational spoken language interface research", Bridging the Gap: Academic and Industrial Research in Dialog Technology workshop at HLT/NAACL 2007

[7] Bohus, Dan & Alexander I. Rudnicky (2009), "The RavenClaw dialog management framework: Architecture and systems", Computer Speech & Language

[8] T. Nestorovic, "General Agent-based Architecture for Collaborative Dialogue Management", in 2nd International Conference on Software Technology and Engineering(ICSTE) 2010, 978-1-4244-8666-3, pp. 207-2011.

[9] A. Raux and M. Eskenazi, "A Multi-layer Architecture for Semi-synchronous Event-driven Dialogue Management", Automatic Speech Recognition & Understanding, 2007. ASRU., Kyoto, pp. 514-519

Optický vláknový refraktometer pre meranie indexu lomu kvapalín

Ľuboš OVSENÍK, Ján TURÁN

Katedra elektroniky a multimediálnych telekomunikácií, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

lubos.ovsenik@tuke.sk, jan.turan@tuke.sk

***Abstrakt* — Tento článok sa zaoberá jednou z oblastí použitia optických vláknových senzorov, a to optickou vláknovou refraktometriou, čo je metóda slúžiaca na meranie malých zmien indexu lomu kvapalín.**

***Kľúčové slová* — meranie indexu lomu kvapalín, optické vláknové senzory, optický vláknový refraktometer**

I. Úvod

Vývoj technológie výroby optických vlákien, vhodných zdrojov a detektorov optického žiarenia a iných vláknových komunikačných systémov, umožňuje v súčasnosti rozsiahle uplatnenie aplikácií týchto systémov pri riešení problémov prenosu informácií. Zdokonaľovaním vlastností optických vláknových komunikačných systémov, klesajúca cena, vzrast spoľahlivosti a životnosti, umožňujú aplikácie optických vláknových komunikačných systémov a vzrastá záujem o optické vlákna aj v neprenosových aplikáciách (vo výskume, v medicíne, v automobilovom a leteckom priemysle, atď.) [1,3]. Pri týchto aplikáciách sa využívajú nielen tie vlastnosti optických vlákien, ktoré sú výhodné pri prenose optických signálov, ale aj tie, ktoré sú nežiaduce a ktoré sa snažíme v komunikačných aplikáciách vylúčiť.

Optické vlákna sa v týchto aplikáciách používajú na osvetľovanie a pozorovanie, v nekonvenčných displejoch, v spínačoch, v systémoch spracovania signálov (pamäte, transverzálne filtre, kódery, dekodéry, korelátory, transformačné procesy, atď.) a v optických vláknových senzoroch rôznych fyzikálnych veličín (posunutia, rýchlosti, tlaku, sily, zrýchlenia, teploty, akustických signálov, intenzity elektrického a magnetického poľa, rádioaktívneho žiarenia, atď.), chemických veličín (PH, koncentrácie) a biologických veličín. Z nekonvenčných aplikácií optických vlákien majú najväčší význam optické vláknové senzory [3,4].

Jednou z oblastí použitia optických vláknových senzorov je aj optická vláknová refraktometria [3,5]. Je to metóda, ktorá slúži na meranie malých zmien indexu lomu kvapalín. Prístroje, ktoré sa na toto meranie využívajú, sa nazývajú optické vláknové refraktometre. Nameraná hodnota indexu lomu kvapaliny nám potom umožní presnejšie definovať vlastnosti meranej kvapaliny, ako je jej hustota, čistota a iné.

II. Optický vláknový refraktometer

A. Rozdelenie refraktometrov

Prvý refraktometer bol skonštruovaný Ernestom Abbé v roku 1869. O päť rokov neskôr 1874 bola publikovaná brožúrka, v ktorej bola popísaná teória merania a samotný refraktometer. Už v roku 1881 bol tento refraktometer priemyselne vyrábaný manufaktúrou Carl Zeiss [2,4].

V súčasnosti sú refraktometre využívané hlavne v potravinárskom, chemickom a farmaceutickom priemysle [2,3].

Pri vyhodnocovaní merania koncentrácie sa okrem indexu lomu používa Brixova stupnica. Brixova stupnica bola pomenovaná po jej tvorcovi - nemeckom vedcovi Adolfovi Brixovi a poskytuje veľkú flexibilitu pre vyhodnocovanie merania. V dnešnej dobe je využívaná v každom priemysle. Je to dodnes najpoužívanejšia stupnica nájdená pre refraktometre.

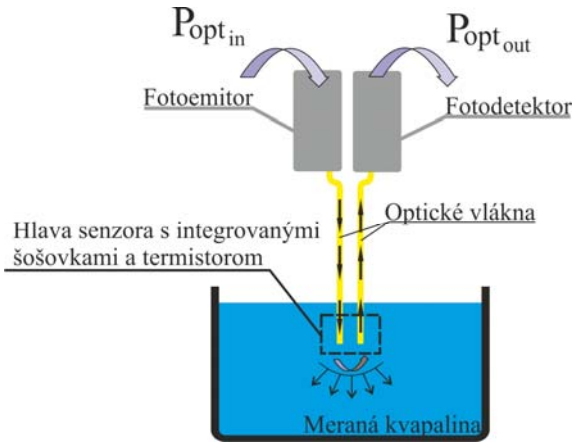
Bezprostredne koreluje so stupnicou indexu lomu *nD*. Je definovaná ako 1g trstinového cukru v 100 ml vody, teda vyjadruje efektívnu hodnotu cukornatosti danej látky. Vo všeobecnosti je Brixova stupnica používaná pre percentuálne vyjadrenie koncentrácie rozpustných tuhých látok vo vode.

Vzhľadom na rôznorodosť použitia metód merania indexu lomu pomocou refraktometrov, potrebou presnosti, či rýchlosti merania by sme tieto mohli rozdeliť do rôznych skupín podľa:

1. Prevodníka (senzora), ktorý využívajú na meranie:
   - refraktometre využívajúce na snímanie optický hranol,
   - refraktometre využívajúce na snímanie optické vlákno.
2. Použitia v praxi:
   - laboratórne refraktometre (ide hlavne o veľmi presné a precízne refraktometre),
   - testovacie refraktometre (zväčša ich predstavujú ručné - handle refraktometre),
   - procesné refraktometre (tzv. In-line refraktometre).
3. Mechanizmu spracovania a vyobrazenia nameraných hodnôt:
   - výlučne optické (využívajúce na vyobrazenie nameranej veličiny len optické prvky),
   - optoelektronické (teda digitálne - spracovanie nameraných hodnôt a vyhodnotenie je zväčša realizované v číslicovej forme, jednoduchšia práca s údajmi).
4. Metódy merania:
   - meranie veľkosti kritického uhla dopadu svetla na definovanej ploche (respektíve porovnáva tmavé a svetlé časti plochy),
   - meranie rozdielu množstva naviazanej a vyžiarenej svetelnej energie do a z optického vlákna, meranie rezonančnej vlnovej dĺžky.

B. Princíp merania indexu lomu látok

Existuje niekoľko spôsobov merania indexu lomu kvapalín. Náš prístroj využíva nasledovný princíp (Obr. 1) [3]. Ak sklenenú tyčinku zbrúsenú na obidvoch koncoch kolmo na svoju os osvetlíme na jednej zo zbrúsených strán, nebude môcť žiarenie vyjsť z tyčinky uloženej vo vzduchu (index lomu vzduchu je podstatne menší než index lomu tyčinky) cez jej plášť, pretože sa od stien odráža vždy pod väčším uhlom než je medzný uhol. Sklená tyčinka teda pôsobí ako vlnovod. Ten istý jav sa bude opakovať aj vtedy, keď tyčinku ohneme do tvaru písmena U. Ak však koleno tyčinky vložíme do kvapaliny s indexom lomu totožným s indexom lomu použitého skla, nebudeme sa na rozhraní obidvoch prostredí žiarenie lámať, ale prejde z tyčinky do kvapaliny. Na druhom konci tyčinky bude nulová intenzita žiarenia. Ak bude index lomu kvapaliny o niečo menší než index lomu skla, bude sa časť žiarenia na rozhraní odrážať späť do tyčinky a časť bude prechádzať do kvapaliny. Na konci tyčinky bude tým väčšia intenzita žiarenia, čím väčší bude rozdiel indexu lomu kvapaliny a tyčinky. Na základe toho možno meraním intenzity žiarenia prejdeného tyčinkou zistiť index lomu skúmanej kvapaliny.



Obr. 1 Princíp refraktometra.

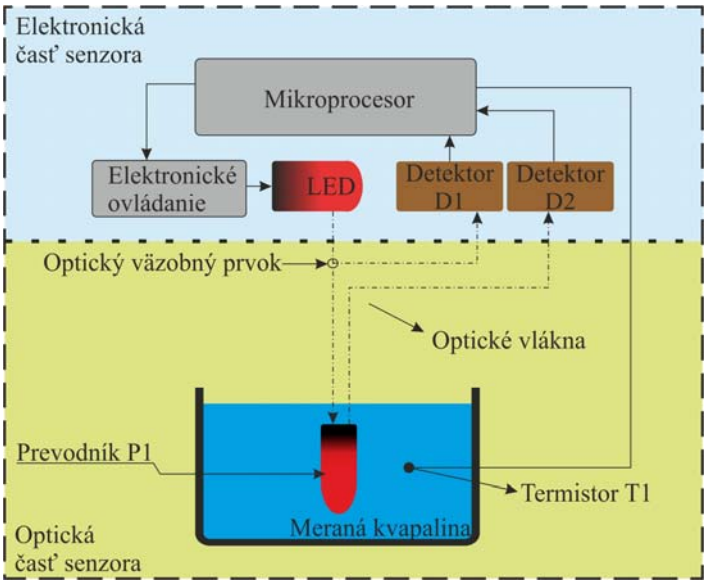
C. Modifikácie refraktometrov

V súčasnosti sa uskutočňuje mnoho výskumných a vývojových prác na rôznych typoch vláknových optických senzoroch, ktoré sa skúšajú v oblastiach ako mechanika, elektromagnetizmus, chémia, atď. [2-5]. Príkladom je inovovaný prevodník indexu lomu, t.j.

refraktometer, ktorý má vynikajúce vlastnosti. Základ senzora pozostáva zo vstupno/výstupného páru mnohovidových vlákien, ktoré sú ukončené malými šošovkami. Kužeľ svetla injektovaný do šošoviek zo vstupného vlákna je odrazený od vnútorného povrchu šošoviek a prenesený späť do výstupného vlákna. Ak vonkajší povrch šošoviek je v kontakte s kvapalinou, potom zníženie intenzity dopadajúceho svetla do výstupného vlákna silne závisí od indexu lomu kvapaliny. Vďaka tomu má prevodník široký dynamický rozsah a jeho relatívna citlivosť pre zodpovedajúcu zmenu indexu lomu Δn je rádu 5 až 10, cez široký rozsah indexu lomu n = 1,3 až 1,6. Priebeh výstupného výkonu v závislosti od indexu lomu môže mať u tohto prevodníka lineárny alebo semilogaritmický charakter. Priemery šošoviek pre prevodník sa môžu pohybovať v rozmeroch od 250 do 350 μm.

Tieto senzory, pre svoju malú štruktúru typu sondy, môžu byť ľahko vložené do vrcholu kontajnera alebo nádoby s kvapalinou. Sú schopné detekovať veľmi malé zmeny indexu lomu. Kombináciou týchto prevodníkov s technológiou číslicového spracovania signálov je možné vyvinúť nové mnohostranné a praktické prístroje. Preto súčasnou snahou je vytvoriť na báze PC riadiaci systém, ktorý riadi laboratórne prístroje tohto typu, ktoré sú schopné určiť index lomu rôznych kvapalín od koncentrácie 1 častica na 10^4 častíc. Ďalšou snahou je vyrobiť riadiaci mikroprocesor a iné typy malých prenosových refraktometrov. Tretím cieľom je vyprodukovať " in line " monitor pre monitorovanie indexu lomu pri použití v chemickom alebo potravinárskom priemysle alebo v iných podobných smeroch spracovania surovín. Refraktometre sú často používané pre medicínske, farmaceutické, chemické, petrochemické, obalové či potravinárske účely, kde s rastúcou koncentráciou určitej látky rastie index lomu meraných kvapalín.

Takmer všetky dostupné refraktometre používajú elementy, v ktorých je umiestnená kvapalná vzorka. Tieto poskytujú výstup, ktorý je založený na stupni úniku svetelného žiarenia na rozhraní kvapalina – optický hranol. V najjednoduchších refraktometroch sa určuje stupeň ohybu vizuálne. V modernejších zariadeniach digitálneho typu je stupeň úniku meraný automaticky, použitím fotodetektora. Pretože index lomu je silne závislý od teploty a tiež v menšej miere od vlnovej dĺžky, tieto efekty musia byť korigované pri návrhu týchto senzorov. Prostriedky, ktoré používa prevodník, musia byť schopné korekcie vnútornej teplotnej závislosti indexu lomu kvapalín. Ďalšie korekcie musia byť urobené tak, aby prevodník pracoval s presnosťou 1 častica na 10^4 častíc. Je nevyhnutné použiť nízkošumové elektronické obvody a korekciu na prúd za tmy fotodetektora, a tiež je nevyhnutné korigovať teplotnú citlivosť svetelného zdroja a fotodetektora aj pre prípadné svetlo pozadia.



Obr. 2 Bloková schéma základného systému optického vláknového refraktometra (lineárna metóda).

Na Obr. 2 je bloková schéma základného typu takéhoto senzora s korigovaním cudzích vplyvov pri meraní. Tento systém pozostáva z nasledujúcich elementov: zdroj svetla (LED alebo LD) a jeho elektronické riadenie; fotodetektor (D1), na presné určenie úrovne výstupného svetla zo zdroja svetla; fotodetektor (D2), na zaznamenanie svetla vracajúceho sa z prevodníka (P1); termistora (T1), na meranie teploty meranej kvapaliny; termistora (T2), na meranie teploty prostredia a mikroprocesora, ktorý riadi tento systém.

V tomto prípade sa bude počítať pomer intenzity z prevodníka pre neznámu kvapalinu a neskôr určenej intenzity svetla pre štandardnú kvapalinu. Mikroprocesor potom výpočtom určí hodnotu indexu neznámej kvapaliny. Teplota okolia a kvapaliny sa určí pomocou termistorov alebo použitím optického vláknového teplotného senzora.

D. Optický vláknový refarktometer ako laboratórny prístroj

Hlavná myšlienka riešenia vychádza z Obr. 2. Celé riešenie vláknového senzorového systému môžeme rozdeliť na tri základné časti:

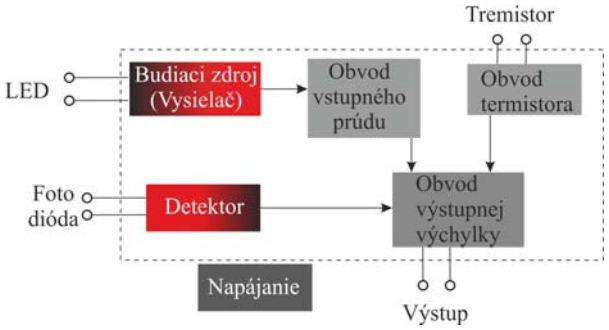
- elektrické obvody pre optický vláknový senzor,
- rozhranie senzor - počítač,
- programový balík pre ovládanie a vyhodnocovanie merania.

Úlohou elektrických obvodov je dodať potrebný napájací prúd pre zdroj svetla (LED, resp. LD) a následne získané napätie z fotodetektora zosilniť a upraviť tak, aby bolo vhodné pre použité rozhranie. Úlohou rozhrania je previesť výstupné analógové napätie zo senzora na číslicovú formu údajov pre počítač. Nakoniec programový balík má zabezpečiť spracovanie údajov a následné zobrazenie výsledkov.

V puzdre optického vláknového senzora je umiestnený termistor pre meranie teploty kvapaliny, ktorej budeme merať zmenu indexu lomu v závislosti od teploty.

Elektrické obvody optického vláknového senzora môžeme rozdeliť na niekoľko častí:

- obvod pre napájanie LED alebo LD - budiaci zdroj,
- obvod pre prijatie a zosilnenie napätia z fotodetektora – detektora,
- obvod pre určenie prúdu, ktorý tečie LED alebo LD,
- obvod termistora,
- obvod pre nastavenie výstupnej výchylky.



Obr. 3 Bloková schéma zapojenia optického vláknového refraktometra ako senzora indexu lomu kvapalín.

Celé zapojenie je schematicky znázornené na Obr. 3. Všetky súčiastky jednotlivých obvodov sú umiestnené na jednej doske plošného spoja, okrem optického vláknového senzora a napájacieho zdroja. Vstupy a výstupy sú zrealizované pomocou BNC konektorov.

1. Budiaci zdroj (vysielač): jeho úlohou je dodávať do LED taký požadovaný napájací prúd, aby sa dosiahla požadovaná intenzita vyžarovaného optického signálu. Má byť schopný plynulo meniť napájací prúd v rozsahu od 0 do 100 mA a je nutné, aby úroveň napájacieho prúdu LED nekolísala.

2. Fotodetektor: jeho úlohou je prijať a následne zosilniť prijatý optický signál na úroveň vhodnú pre ďalšie spracovanie. Fotodióda pretransformuje intenzitu svetla na napätie. Fotodióda môže pracovať v dvoch módoch - fotovoltickom a fotokonduktívnom. Vo fotovoltickom móde pracuje fotodióda nezávisle od vonkajšieho zdroja, pričom pri fotokonduktívnom móde napätie na fotodióde závisí od vonkajšieho záverného napájania. Pri fotovoltickom móde je prúd za tmy minimálny.

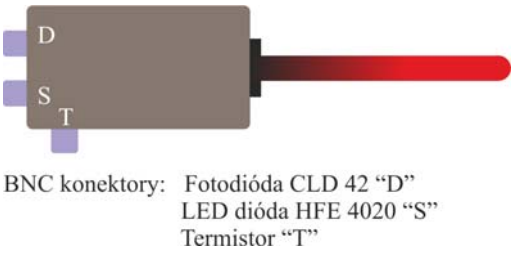
3. Indikátor vstupného prúdu: jeho funkciou je určiť prúd tečúci LED diódou. Princíp spočíva v meraní napätia na známom odpore. Výsledný prúd sa získa dosadením hodnôt do Ohmovho vzťahu ($I = U/R$).

4. Obvod výstupnej výchylky: pomocou tohto obvodu je možné nastaviť výstupné napätie tak, aby jeho zmena odpovedala indexu lomu danej kvapaliny alebo rozdielu indexov lomu dvoch porovnávaných kvapalín.

5. Napájanie: je zabezpečené symetrickým zdrojom ±15 V.

6. Optický vláknový senzorový modul (napr. z LED): tvorí ho kovové puzdro z nehrdzavejúcej ocele (Obr. 4). Na konci vyvedenej rúrky je otvor pre výstup optického vlákna.

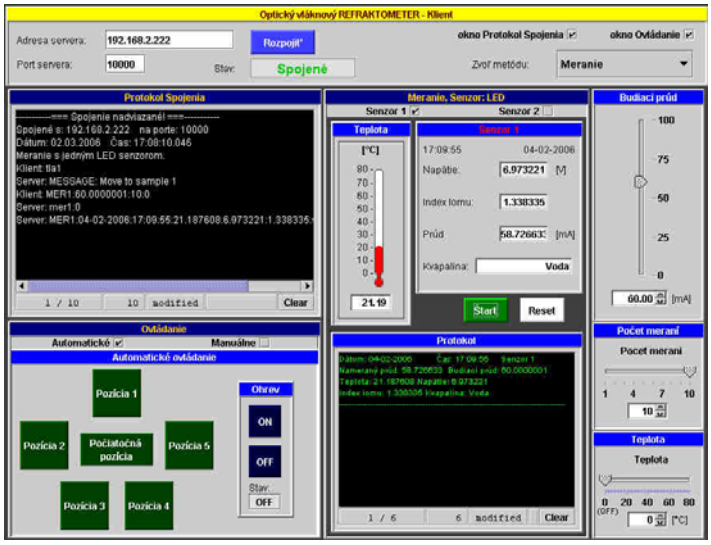
Vo vnútri puzdra sú umiestnené LED, fotodióda a termistor pre meranie teploty meranej kvapaliny. Spojenie senzora s elektronikou je realizované pomocou tienených vodičov a BNC konektorov.



Obr. 4 Puzdro s optickým vláknovým senzorovým modulom.

E. Programové vybavenie refarktometra

Program pre prácu a riadenie meracieho pracoviska s laboratórnym refraktometrom je vytvorený v prostredí LabWindows/CVI od firmy National Instruments pre operačné systémy Windows. LabWindows/CVI je ANSI C vývojové prostredie špecializované pre návrh meracích, testovacích a riadiacich aplikácií. Obsahuje množstvo knižníc pre riadenie zariadení, zber dát, analýzu a užívateľské rozhranie.



Obr. 5 Ponuka hlavného okna pre „Meranie" optickým vláknovým refraktometrom.

Ponuka hlavného menu (Obr. 5) je rozdelená do viacerých častí – okien. Okno v ľavom dolnom rohu „Ovládanie" je určené pre zvolenie si kvapaliny, ktorú chceme merať. Je v ňom možné si zvoliť buď automatické alebo manuálne ovládanie. „Pozícia 1,2,3,4,5" predstavuje kvapaliny v zásobníku, ktoré si môžeme vybrať na meranie a môžeme sa rozhodnúť pre ich prípadný ohrev „Ohrev ON resp. OFF"

V ďalšom okne „Meranie, Senzor" si môžeme navoliť druh merania a typ senzora. Buď meriame jednoduchou metódou LED alebo LD, alebo si zvolíme komparatívnu metódu, teda použitie dvoch senzorov LED alebo LD súčasne. Podľa toho sa nám nastaví rozsah stupnice v okne „Budiaci prúd".

Meranie uskutočňujeme tak, že nastavíme „Budiaci prúd", „Štatistický počet meraní" a nami zvolenú „Teplotu" pri akej chceme kvapalinu merať. Tlačidlo „Štart" slúži na spustenie merania. V pravej spodnej časti appletu, sa nachádza okno so samotnými výsledkami meraní „Protokol". Všetky staršie merania sa zobrazujú v okne „Záznam". Na vymazanie hodnôt slúži tlačidlo „Reset".

III. EXPERIMENTY S OPTICKÝM VLÁKNOVÝM REFRAKTOMETROM

Na experimentálne overenie správnosti fungovania optického vláknového refraktometra sme použili meranie indexu lomu rôznych kvapalín. Meranie bolo vykonávané lineárnou

(jednoduchou, jedno-senzorovou) metódou pomocou rôznych typov senzorových modulov (LED, LD). Výsledky sú v Tabuľke 1.

Ako vyplýva z výsledkov výpočtov, chyby merania indexu lomu neznámych kvapalín sa pohybujú do 1%. Výsledky týchto experimentálnych meraní, ktoré boli vykonané na oboch typoch senzorov, dokazujú dobrú zhodu s teoretickými predpokladmi pre daný druh merania. Z toho vyplýva, že navrhnuté obvodové riešenie optických vláknových refraktometrov je správne a funkčné. Odchýlky, ktoré vznikajú šumom v analógovej časti, sú čiastočne eliminované štatistickými metódami v programe. Senzor LD vykazuje väčšiu presnosť, pretože obsahuje koherentný zdroj žiarenia, čo robí celý systém stabilnejší.

Tabuľka 1
Namerané hodnoty kvapalín senzorom LED a LD

Merané kvapaliny	Index lomu n_i (ideal.)	Senzor LED				Senzor LD			
		Detek. Napätie (V)	Pomer bud./det.	Index lomu n_n (namer.)	Chyba $n_n/n_i.100$ (%)	Detek. Napätie (V)	pomer bud./det.	Index lomu n_n (namer.)	Chyba $n_n/n_i.100$ (%)
Voda	1,333	9,6	0,52	1,3335	0,03	4,14	0,38	1,3332	0,015
Lieh	1,362	4,2	1,19	1,3680	0,40	2,10	0,75	1,3601	0,13
Glycol	1,4268	0,84	5,952	1,4295	0,18	0,83	1,90	1,4271	0,02
MOBIL VS200	1,4399	0,62	8,064	1,4421	0,15	0,75	2,102	1,4425	0,18
MADIT-prevod	1,4828	0,08	62,5	1,4830	0,01	0,50	3,154	1,4835	0,04

IV. ZÁVER A ZHODNOTENIE

Optický vláknový refraktometer umožňuje meranie indexu lomu nielen lineárnou metódou, t.j. jedným senzorom (LED alebo LD) v dvoch kanáloch, ale aj diferenčnou metódou, t.j. s použitím dvoch identických senzorov (LED alebo LD). Je to možné preto, lebo v prístroji sú zabudované dve rovnocenné analógové časti a je len vecou zapojenia optických vláknových senzorových modulov a programového vybavenia, ktorý spôsob merania sa bude využívať. Ak sa užívateľ rozhodne pre priamu metódu, stačí k meraniu jeden senzorový modul ľubovoľného typu. Počítač potom pomocou svojho programového vybavenia určí hodnotu indexu neznámej kvapaliny z pomeru intenzity optického signálu z prevodníka pre neznámu kvapalinu a z vopred určenej intenzity svetla pre štandardnú kvapalinu. Ak sa užívateľ rozhodne pre diferenčnú metódu, sú potrebné k meraniu dva identické senzorové moduly. V tomto prípade sa indexy lomu známej a neznámej kvapaliny merajú súčasne. Táto metóda môže byť použitá, ak sa vyžaduje extrémna presnosť.

POĎAKOVANIE

Podporujeme výskumné aktivity na Slovensku/ Projekt je spolufinancovaný zo zdrojov EÚ. Tento článok bol vypracovaný v rámci projektu "Centrum excelentnosti integrovaného výskumu a využitia progresívnych materiálov a technológií v oblasti automobilovej elektroniky", ITMS 26220120055 (50%). Táto práca bola taktiež podporovaná a realizovaná z projektu Rozvoj Centra informačných a komunikačných technológií pre znalostné systémy (ITMS kód: 26220120030) na základe podpory operačného programu Výskum a vývoj financovaného z Európskeho fondu regionálneho rozvoja (50%).

LITERATÚRA

[1] V. G. Gomes, et al., „Web-based Courseware in Teaching Laboratory-based Course“, *Global J. of Eng. Education*, vol.4, no.1, 2002, pp. 65-71.

[2] N. Kapany, J. Pike, *Automatic correction of residual aberrations: Fiber Optics*. New Yourk, December 1957, pp. 1109 – 1117.

[3] J. Turán, Ľ. Ovseník, *Projektovanie optických vláknových komunikačných a senzorových systémov*. Košice, Harlequin, 2007.

[4] J. Turán, Ľ. Ovseník, J. Turán, Jr., J. Vásárhelyi, „WWW Based Remote Control Application in Optoelectronic Systems“, *Carpathian Journal of Electronic and Computer Engineering*, vol.1, no.1, 2008, pp. 69-75.

[5] J. Turán, Ľ. Ovseník, J. Turán, Jr., „Visualisation in Web Controlled FiberOptic Refractometer System“, *MIPRO 2008: 31st International Convention*, Opatija, Croatia, May 26-30, 2008, pp. 100-103.

Optický vláknový refraktometer riadený cez internet

Ľuboš OVSENÍK, Ján TURÁN

Katedra elektroniky a multimediálnych telekomunikácií, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

lubos.ovsenik@tuke.sk, jan.turan@tuke.sk

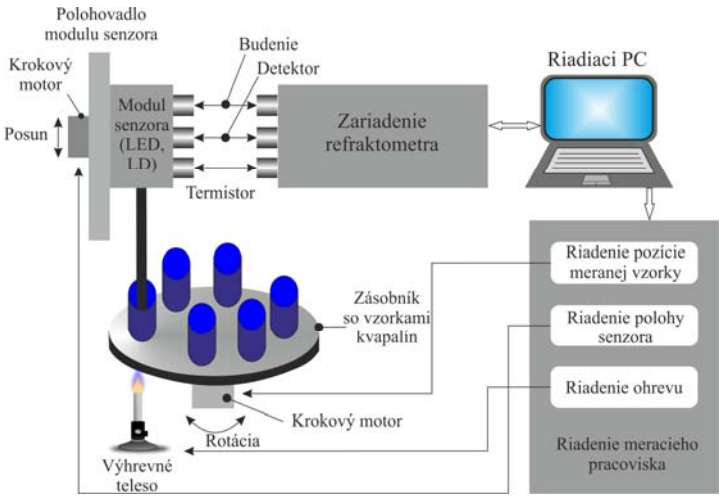
***Abstrakt* — Tento článok sa zaoberá optickým vláknovým refraktometrom pre meranie malých zmien indexu lomu kvapalín, ktorý je súčasťou WWW (web) riadeného laboratória so vzdialeným prístupom.**

***Kľúčové slová* — laboratórium so vzdialeným prístupom, meranie indexu lomu kvapalín, optický vláknový refraktometer, Web-riadené laboratórium**

I. Úvod

Základným cieľom WWW (web) riadených laboratórií so vzdialeným prístupom je umožniť študentovi vykonávať experiment bez fyzickej prítomnosti priamo v experimentálnom laboratóriu [1,3]. Hlavnými úlohami cez Web-riadených laboratórií so vzdialeným prístupom je umožniť:

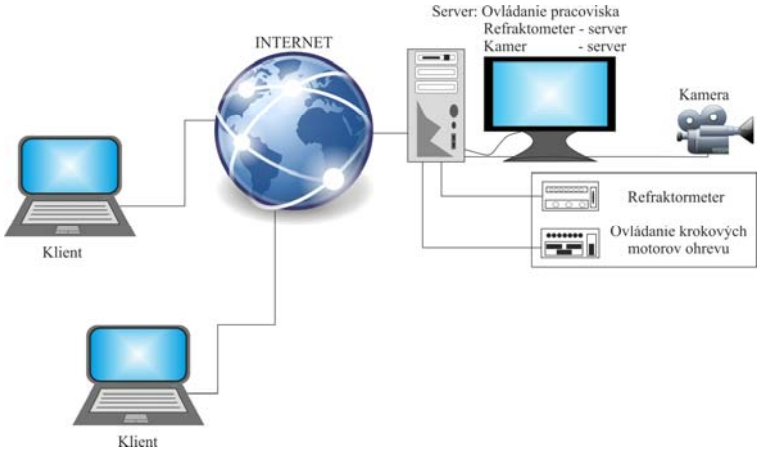
- Tréning študentov – začiatočníkov;
- Opakovanie merania a overovanie výsledkov merania mimo času priamej výučby v laboratóriu;
- Maximálne časové využitie a vyťaženie prístrojového vybavenia vzdelávacej inštitúcie bez potreby neustálej prítomnosti administrátora v laboratóriu a tým zmenšenie prevádzkových nákladov laboratória;
- Vykonávať nebezpečné merania bez ohrozenia študenta (vysoké napätia, rádioaktívne žiarenie, chemické a biologické ohrozenie atď.);
- Vykonať experiment v laboratóriu s unikátnym vybavením na inej vzdialenej univerzite alebo inštitúcii, kde by fyzická prítomnosť študenta vyžadovala neúmerné finančné náklady (cestovné, ubytovanie atď.).



Obr. 1 Všeobecná hardvérová infraštruktúra meracieho pracoviska.

Spojenie technológií web riadených laboratórií so vzdialeným prístupom s multimediálnymi nástrojmi a modernými komunikačnými technológiami v systémoch elektronického vzdelávania umožňujú odľahčiť učiteľa najmä v počte fyzicky odučených hodín, súčasne s prehĺbením

teoretických a praktických znalostí získavaných študentom – zvyšuje efektivitu vzdelávacieho procesu [3]. Príkladom takéhoto laboratória je aj nami navrhnuté a zrealizované meracie pracovisko [4,5], ktorého cieľom je dosiahnuť automatizáciu pri meraní indexu lomu kvapalín s laboratórnym optickým vláknovým refraktometrom (Obr. 1). Toto tele – pracovisko umožňuje vzdialeným skupinám užívateľov podieľať sa spoločne na aplikovaní laboratórnych meraní, pokiaľ súčasne komunikujú a používajú výpočtové zdroje. Využívajú sa aplikácie, ktoré komunikujú s hardvérom prostredníctvom internetu. Využíva sa komunikácia medzi klientom a serverom nazývaná Klient – Server aplikácia. Všeobecná hardvérová infraštruktúra zariadenia je na Obr. 2. Základnú časť tvoria server a Riadiaci PC. Server poskytuje *html* stránku a taktiež komunikáciu medzi klientom a Riadiacim PC, ku ktorému je pripojené zariadenie na ovládanie refraktometra a automatizácie merania.



Obr. 2 Hardvérová štruktúra vzdialeného meracieho pracoviska s refraktometrom.

II. AUTOMATIZOVANÉ MERACIE PRACOVISKO OVLÁDANÉ CEZ WWW

A. Hardvérová infraštruktúra a softvérová architektúra

Komunikácia medzi vzdialeným užívateľom (klientom) a počítačom, poskytujúcim ovládanie pracoviska (serverom) sa zakladá na architektúre klient–server. Architektúra klient–server vychádza z myšlienky prezentácie dát spracovaných serverom, napr. na strane klienta sa zobrazujú postupne prebiehajúce výpočty na výkonnom servery. Klientom môže byť tiež prehliadač internetových stránok, ktorý využíva výpočtovú a databázovú silu vzdialeného uzla siete k zobrazeniu časti serverom poskytovaných informácií. Klient posiela svoje požiadavky na server, ktorý ich spracuje a po vykonaní požadovanej činnosti posiela odpoveď späť klientovi. Z toho vyplýva jedna zo základných myšlienok tejto architektúry, a to minimálna hardvérová ako aj softvérová náročnosť zo strany klienta.

V našom prípade server vykonáva nasledujúce úlohy:

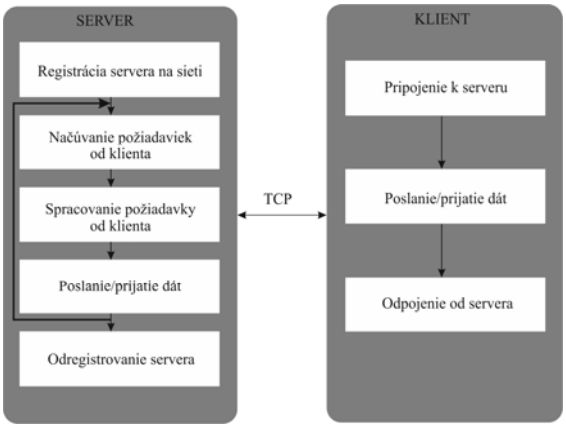
- refraktometer–server: prijíma a po prijatí spracováva požiadavky od klienta na pripojenie ku serveru, vykonanie požadovanej obsluhy meracieho pracoviska a taktiež posiela klientovi odpovede,
- ovládanie meracieho pracoviska, t.j. riadenie jednotlivých funkcií refraktometra, otáčanie vzoriek a posun senzora, na základe požiadaviek klienta,
- obrazové spojenie na sledovanie správnosti činnosti meracieho pracoviska pomocou web kamery. Túto úlohu zabezpečuje kamera–server, ktorý po prijatí požiadavky na obrazové spojenie vysiela video po sieti (video streaming) ku klientovi.

Úlohou klientskej časti aplikácie je:

- pripojenie ku serveru,
- ovládanie meracieho pracoviska a nastavovanie parametrov pre meranie, riadenie merania,
- zobrazovanie výsledkov vo vhodnom grafickom rozhraní,
- zabezpečenie obrazového spojenia, klient ku kamera–server, t.j. pripojenie ku kamera–server a zobrazenie prijímaného videa.

B. TCP/IP protokoly, TCP sieťové aplikácie

Počítače v počítačových sieťach používajú pre vzájomnú komunikáciu sieťové protokoly. Existuje celá rada sieťových protokolov. V Internete a v mnohých iných lokálnych počítačových sieťach sa používajú sieťové protokoly TCP/IP. TCP/IP (Transmission Control Protocol/Internet Protocol) je súbor protokolov (pravidiel), ktoré popisujú informačný tok medzi počítačmi zapojenými do spoločnej siete. TCP (patrí do transportnej vrstvy) zriadi spojenie pre prenos dát a IP (patrí do sieťovej vrstvy) definuje metódu pre posielanie dát. Typická sieťová komunikácia používajúca TCP zahŕňa klient–server architektúru (Obr. 3).



Obr. 3 TCP komunikácia.

Klient aj server sú na sieti jednoznačne identifikovaní svojou IP adresou. Pred vytvorením TCP sieťového spojenia serverovská aplikácia musí byť registrovaná na sieti. Registráciou sa zriadi sieťový port, cez ktorý klientská aplikácia pristupuje na server. Po registrácií serverová aplikácia na danom sieťovom porte načúva pre prichádzajúce požiadavky od klienta. Aby sa klient vedel pripojiť ku serveru, musí poznať meno alebo sieťovú adresu host-a, na ktorom serverová aplikácia beží a číslo portu, na ktorom načúva. Server po prijatí žiadosti od klienta, žiadosť spracuje a pošle odpoveď klientovi. Vzájomné pôsobenie medzi TCP serverom a TCP klientskou aplikáciou je znázornené na Obr. 3. Keďže TCP je spojovo – založený protokol, tak server a klient musia najprv vytvoriť spojenie medzi sebou a až potom si môžu vymieňať dáta.

C. Komunikácia medzi klientom a serverom

Komunikácia medzi serverom (refraktometer-server) a klientom (refraktometer-klient) prebieha prostredníctvom protokolu TCP/IP. Po vytvorení spojenia klient a server komunikujú tak, že si posielajú správy obsahujúce textové reťazce, ktoré odpovedajú kladeným požiadavkám zo strany klienta a odpovediam na tieto požiadavky zo strany servera [3-5].

Obsah textových reťazcov zodpovedá formátu, ktorý sme navrhli. Podľa tohto formátu sa textové reťazce vytvárajú pre žiadosť zo strany klienta aj pre odpoveď zo strany servera. Po prijatí správy sa podľa neho spracúva textový reťazec, ktorý prijatá správa obsahuje. Správy s informatívnym charakterom majú svoj vlastný formát. Rozlišujeme dva typy týchto správ, a to hlásenia priebehu (MESSAGE) a chybové hlásenia (ERROR). Po označení typu nasleduje dvojbodka a výstižný popis udalosti. Obidva typy správ sú posielane len zo strany servera a majú za úlohu informovať klienta o priebehu, prípadne chybách pri riadení meracieho pracoviska. Správy nie sú nijako spracovávané, ale rovno zobrazované klientovi v grafickom rozhraní na to určenom. Vytváranie reťazca pre požiadavku zo strany klienta a jeho poslanie serveru po zvolení činnosti a nastavení požadovaných parametrov prebieha podľa algoritmu:

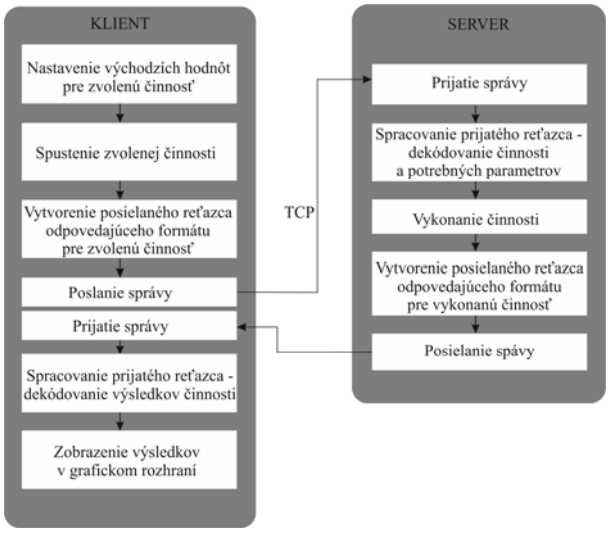
- načítanie parametrov pre zvolenú činnosť,
- vytvorenie posielaného reťazca podľa formátu zvolenej činnosti,
- poslanie správy.

Po prijatí správy serverom nasleduje spracovanie reťazca:

- z reťazca sa zisťuje činnosť, ktorá sa má vykonať,
- podľa formátu pre zistenú činnosť sa dekódujú z reťazca parametre pre požadovanú činnosť,
- vykonanie činnosti.

Po vykonaní činnosti server vygeneruje podľa formátu pre ukončenú činnosť posielaný reťazec tým istým postupom ako klient pre požiadavku s tým rozdielom, že parametre reťazca

sú výsledky činnosti. Po prijatí odpovede od servera klient vykonáva spracovanie reťazca postupom ako server. Dekódované parametre z reťazca (výsledky činnosti) potom zobrazuje v grafickom rozhraní na to určenom. Tento postup je znázornený na nasledujúcom obrázku (Obr. 4). Pri prijatí nedefinovanej správy (správa bola pri prenose poškodená) blok spracovania prijatej správy túto správu nevie dekódovať. Zo strany servera v tomto prípade dôjde k poslaniu upozorňujúcej správy o tejto skutočnosti klientovi, ktorá mu je zobrazená, v prípade prijatia nedefinovanej správy klientom je klient o tom informovaný príslušným hlásením.



Obr. 4 Postup vytvárania posielaných a spracovania prijatých textových reťazcov.

V nami definovanej komunikácií existuje ešte jeden typ správy. Správa obsahuje textový reťazec LIVING a slúži na zistenie živosti spojenia. Posielanie začína zo strany klienta a ten v 30 sekundových intervaloch sleduje, či mu v tomto limite prišla odpoveď zo strany servera na túto požiadavku. Ak nie, považuje spojenie za prerušené a upozorní o tejto skutočnosti užívateľa. Ak mu odpoveď prišla, znovu pošle správu LIVING a opakuje uvedený postup. Na strane servera to prebieha podobne. Server po prijatí LIVING odpovie príslušnou správou a v 30 sekundových intervaloch zisťuje či mu prichádza správa LIVING. Ak nie, považuje spojenie za prerušené, v opačnom prípade sa celý postup opakuje.

D. Video streaming

Video streaming je pojem pre označenie prenosu videa po počítačovej sieti. Pri streamovaní koncový užívateľ sleduje video okamžite od pripojenia a nie je nutné si video najprv uložiť (download). Tieto výhody streamovania (okamžité sledovanie videa) umožňujú využiť tieto techniky pri živých vysielaniach (live broadcasting) niekedy označované ako webcast (webové vysielanie) alebo netcast. Pre sledovanie streamovaného videa je potrebný špeciálny streamovací server, ktorý vytvorí (encode) požadovaný dátový tok (stream) pre prenos. Klient prijíma stream, spracuje (decode) ho a zobrazí užívateľovi. Pri streamovaní je teda dôležitý formát dátového toku (kontajner dátového toku) a použitý sieťový protokol pre prenos.

My sme použili na streamovanie videa (kamera – server) ako aj klienta už existujúci open source program VLC media player. Z viacerých možností, ktoré VLC media player ponúkal sme pre streamovanie videa zvolili nasledovné kroky:

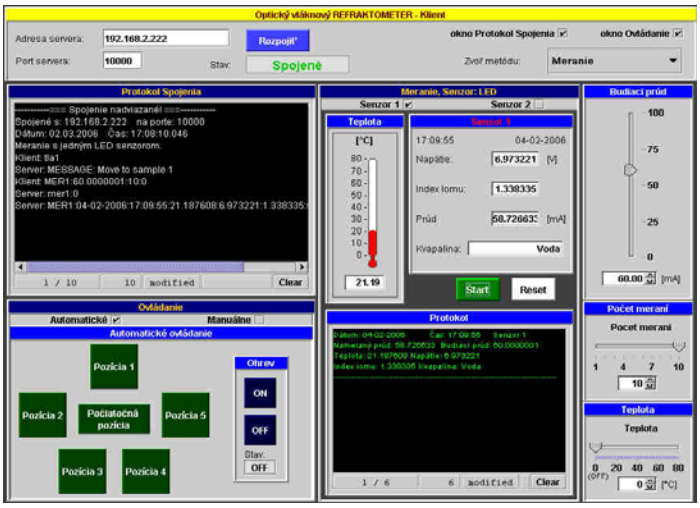
1. na strane streamovacieho servera beží jednoduchý Http server, ktorý na zadanom porte načúva pre požiadavku klienta o streamovanie videa;
2. po prijatí požiadavky od klienta sa vytvorí TCP/IP spojenie na zadanom porte a začne sa streamovať video ku klientovi.

Pre formát videa z web kamery sme zvolili formát MPEG-4 s dátovým tokom 256kbit/s a pre formát dátového toku streamovaného videa kontajner Ogg. Ogg (Operation Good Guys) je bitstream kontajner navrhnutý pre efektívne streamovanie. Ogg bitstream kontajner bol vytvorený pre rozvoj kódovania a dekódovania multimédií, pričom je voľne dostupný a môže sa voľne implementovať do iných softvérov.

E. Programové vybavenie refarktometra pre časť - Klient

Program pre prácu a riadenie meracieho pracoviska s laboratórnym refraktometrom je

vytvorený v prostredí LabWindows/CVI od firmy National Instruments pre operačné systémy Windows. Po spustení programu je nutné aby sa klient pripojil na server, na ktorom je pripojený refraktometer. To sa urobí vpísaním príslušnej IP adresy a portu servera v ľavej hornej časti hlavného okna programu (Obr. 5). Všetky príkazy riadenia meracieho pracoviska sú aktívne až po pripojení sa na server o čom informuje „Stav: Spojené“. Potom môže klient začať so samotným meraním.



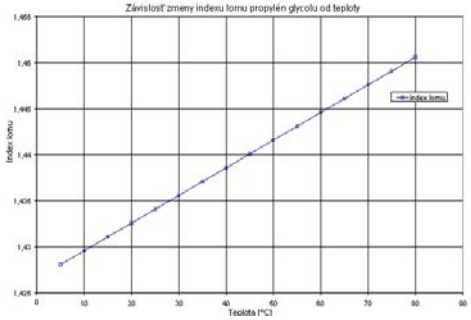
Obr. 5 Ponuka hlavného okna pre „Meranie“ optickým vláknovým refraktometrom.

Ponuka hlavného menu (Obr. 5) je rozdelená do viacerých častí – okien. Okno v ľavom dolnom rohu „Ovládanie“ je určené pre zvolenie si kvapaliny, ktorú chceme merať. Je v ňom možné si zvoliť buď automatické alebo manuálne ovládanie. „Pozícia 1,2,3,4,5“ predstavuje kvapaliny v zásobníku, ktoré si môžeme vybrať na meranie a môžeme sa rozhodnúť pre ich prípadný ohrev „Ohrev ON resp. OFF“ V ďalšom okne „Meranie, Senzor“ si môžeme navoliť druh merania a typ senzora. Buď meriame jednoduchou metódou LED alebo LD, alebo si zvolíme komparatívnu metódu, teda použitie dvoch senzorov LED alebo LD súčasne. Podľa toho sa nám nastaví rozsah stupnice v okne „Budiaci prúd“. Meranie uskutočňujeme tak, že nastavíme „Budiaci prúd“, „Štatistický počet meraní“ a nami zvolenú „Teplotu“ pri akej chceme kvapalinu merať. Tlačidlo „Štart“ slúži na spustenie merania. V pravej spodnej časti appletu, sa nachádza okno so samotnými výsledkami meraní „Protokol“. Všetky staršie merania sa zobrazujú v okne „Záznam“. Na vymazanie hodnôt slúži tlačidlo „Reset“.

III. Experimenty s optickým vláknovým refraktometrom

A. Zmena indexu lomu kvapalín v závislosti od teploty

V tomto experimente bola overená činnosť prístroja pri meraní zmeny indexu lomu predloženej kvapaliny v závislosti od zmeny teploty. Pre meranie bola použitá kvapalina propylén glycol. Je to bezfarebná tekutina bez vône, sladkej chuti, ktorá je miešateľná s vodou a s väčšinou organických rozpúšťadiel. Index lomu propylén glycolu je pre sodíkové svetlo 1,4326 pri 20 ºC. Pre meranie teplotnej závislosti zmeny indexu lomu propylén glycolu bola menená teplota od 5 do 80 ºC s krokom 5 ºC. Výsledky merania sú na Obr. 6.



Obr. 6 Závislosť zmeny indexu lomu od teploty pre propylén glycol.

Z nameraných hodnôt vyplýva, že teplotná závislosť zmeny indexu lomu propylén glycolu $\Delta n/\Delta t$ je lineárna a mení sa s rýchlosťou 0,0003/ºC.

Ďalšími teplotnými závislosťami kvapalín, na ktorých sme overovali funkčnosť systému boli: voda, nízkoenergetický sirup, káva, prevodový a motorový olej. Merania prevodového a motorového oleja boli uskutočnené pri teplotách od 20 °C do 80 °C, odstupňované po 10 °C a merania vody, nízkoenergetického sirupu a kávy boli uskutočnené pri teplotách od 20 °C do 50 °C, odstupňované po 5 °C.

Z nameraných hodnôt pre vodu, kávu a sirup vyplýva, že s narastajúcou teplotou narastá aj index lomu. Pri zohratí vody o 1 °C sa index približne zvýši o 0,0001, pri káve o 0,00048, pre sirup je o 0,0004 väčší na 1 °C.

Pre motorový olej sa zvyšovaním teploty index lomu zmenšuje. Pokles indexu lomu na 1 °C je 0,00007. To isté platí aj pre prevodový olej, kde index lomu na 1 °C je o 0,00031 menší.

B. Zmena indexu lomu kvapalín v závislosti od koncentrácie

V ďalšom experimente bola overená činnosť prístroja pri meraní zmeny indexu lomu predloženej kvapaliny v závislosti od veľkosti koncentrácie. Pre meranie bola použitá kvapalina propylén glycol, ktorý bol rozpúšťaný v destilovanej vode od koncentrácie 0 % do koncentrácie 100 %, pri konštantnej laboratórnej teplote 25 ºC.

Pre zmenu indexu lomu propylén glycolu v závislosti od jeho koncentrácie v destilovanej vode platí lineárna závislosť tejto zmeny $\Delta n/\Delta c$ s krokom približne 0,001/1%.

IV. ZÁVER A ZHODNOTENIE

Optický vláknový refraktometer ovládaný cez internet komunikuje s počítačom pomocou ovládacieho programového vybavenia, vytvoreného v programovacom prostredí Java a LabWindows/ CVI, cez paralelné rozhranie počítača. Analógová časť, spustenie prevodu, ako aj načítanie odmeraných vzoriek sú riadené počítačom, pomocou vytvoreného programu. Program je navrhnutý s cieľom jednoduchej obsluhy a ovládania refraktometra tak, že umožňuje merať indexy lomu kvapalín od 1,3 až po 1,6, testovať LED a LD senzory s budením do 100 mA a tiež načítať údaje z prenosného refraktometra cez paralelné rozhranie RS232C. Program v spolupráci s prístrojom namerané hodnoty spracuje a po spracovaní vypisuje vo výstupnom okne.

POĎAKOVANIE

Podporujeme výskumné aktivity na Slovensku/ Projekt je spolufinancovaný zo zdrojov EÚ. Tento článok bol vypracovaný v rámci projektu "Centrum excelentnosti integrovaného výskumu a využitia progresívnych materiálov a technológií v oblasti automobilovej elektroniky", ITMS 26220120055 (50%). Táto práca bola taktiež podporovaná a realizovaná z projektu Rozvoj Centra informačných a komunikačných technológií pre znalostné systémy (ITMS kód: 26220120030) na základe podpory operačného programu Výskum a vývoj financovaného z Európskeho fondu regionálneho rozvoja (50%).

LITERATÚRA

[1] V. G. Gomes, et al., „Web-based Courseware in Teaching Laboratory-based Course“, *Global J. of Eng. Education*, vol.4, no.1, 2002, pp. 65-71.

[2] N. Kapany, J. Pike, *Automatic correction of residual aberrations: Fiber Optics*. New Yourk, December 1957, pp. 1109 – 1117.

[3] J. Turán, Ľ. Ovseník, *Projektovanie optických vláknových komunikačných a senzorových systémov*. Košice, Harlequin, 2007.

[4] J. Turán, Ľ. Ovseník, J. Turán, Jr., J. Vásárhelyi, „WWW Based Remote Control Application in Optoelectronic Systems“, *Carpathian Journal of Electronic and Computer Engineering*, vol.1, no.1, 2008, pp. 69-75.

[5] J. Turán, Ľ. Ovseník, J. Turán, Jr., „Visualisation in Web Controlled FiberOptic Refractometer System“, *MIPRO 2008: 31st International Convention*, Opatija, Croatia, May 26-30, 2008, pp. 100-103.

Analógový pásmový ekvalizér s digitálnym riadením

Marek PETRINEC, Miloš DRUTAROVSKÝ

Katedra elektroniky a multimediálnych telekomunikácií, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

marek.petrinec@student.tuke.sk, milos.drutarovsky@tuke.sk

***Abstrakt* — Príspevok opisuje základné hardvérové a softvérové bloky analógového pásmového ekvalizéra s digitálnym riadením, ktorý v analógovej časti využíva pásmové priepusty tvorené fázovacími článkami a rozdielovým zosilňovačom. Ako riadiaci obvod je použitý procesor STM32F107 s jadrom ARM Cortex-M3. Rozhranie analógovej a digitálnej časti je tvorené digitálnymi potenciometrami DS1882 od firmy Maxim.**

***Kľúčové slová* — digitálne riadenie, ekvalizér, fázovací článok, I2C zbernica, pásmový priepust**

I. Úvod

Ekvalizér je jednou z najdôležitejších častí ozvučovacej sústavy. Umožňuje kompenzovať negatívne akustické vlastnosti priestoru v ktorom je signál reprodukovaný a umožňuje tiež kompenzovať niektoré nedostatky zdroja akustického signálu. Ekvalizéry majú široké uplatnenie vo všetkých zvukových zariadeniach a preto boli vyvinuté rôzne druhy ekvalizérov ako napríklad ekvalizéry parametrické alebo grafické [1].

V bežnej amatérskej tvorbe analógových ekvalizérov sa najčastejšie využívajú zapojenia aktívnych pásmových priepustov (BP – Band-Pass filter). Tieto filtre ale vytvárajú fázový posun ktorý je často rozdielny na blízkych frekvenciách, čo spôsobuje výrazné skreslenie signálov týchto frekvencií pri reprodukcii audio signálov. Tento jav je v audio technike veľmi nežiaduci.

Úpravy zvuku môžu vykonávať aj mikroprocesory alebo špecializované signálové procesory, ktoré spracúvajú postupnosť digitálnych vzoriek. Tieto vzorky sú väčšinou získavané z A/D prevodníkov v prípade že je vstupný signál analógový alebo sú privádzane cez impedančný prispôsobovač priamo na vstup obvodu v prípade, že vstup je digitálny. Vtedy hovoríme že ekvalizér je digitálny.

Ekvalizéry sa delia aj podľa strmosti útlmu v okolí hraničnej frekvencie, ktorú vyjadruje koeficient Q použitého filtra. Podľa strmosti útlmu amplitúdy hovoríme o strmosti 6dB/okt (filter I. stupňa), strmosti 12dB/okt (filter II. stupňa), strmosti 18dB/okt (filter III. stupňa), atď. So stupňom použitého filtra stúpa aj zložitosť zapojenia analógových ekvalizérov. Digitálne ekvalizéry majú preladiteľnú strmosť bez nutnosti použitia zložitejšieho vnútorného zapojenia, keďže všetky výpočty sú realizované číslicovým signálovým procesorom. V prípade že strmosť je nastaviteľná hovoríme o parametrických ekvalizéroch. V opačnom prípade sa tieto ekvalizéry nazývajú grafické.

V ďalšom texte opíšeme iba základné analógové a digitálne časti ekvalizéra, ktorý vznikol v rámci bakalárskej práce [11], ktorá obsahuje podrobné schémy zapojení, predlohy plošných spojov a opis riadiaceho softvéru.

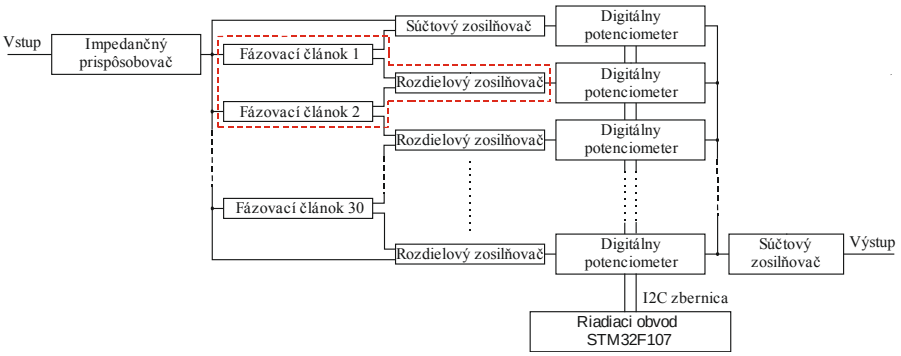
II. Blokova štruktúra ekvalizéra

Bloková schéma navrhnutého ekvalizéra je zobrazená na Obr.1 a je tvorená analógovou časťou, ktorá je riadená riadiacim procesorom.

A. Analógová časť

Základ analógovej časti tvorený množinou navzájom previazaných pásmových priepustov s využitím fázovacích článkov bol prevzatý z [2]. Vstupný analógový signál vstupuje do

impedančného prispôsobovača, ktorý oddeľuje vstupný signál od samotných fázovacích článkov. Toto zapojenie zabraňuje preťaženiu výstupných obvodov zdroja signálu a umožňuje regulovať vstupnú citlivosť ekvalizéra. Za impedančným prispôsobovačom nasledujú fázovacie články. Tieto články majú prenosovú funkciu rovnú jednej avšak realizujú frekvenčne závislý fázový posun v závislosti na parametroch (R, C prvkoch) použitých vo fázovacích článkoch.



Obr. 1 Bloková schéma ekvalizéra s vyznačeným blokom jedného pásmového priepustu.

Každý fázovací článok je nastavený na odlišnú medznú frekvenciu (definovanú ako frekvencia, pri ktorej dochádza k fázovému posunu 90 stupňov), čím dochádza k rozdielnemu posuvu fáz signálov, ktoré sú privedené na vstupy každého rozdielového zosilňovača. Ten na svoj výstup prenesie rozdiel týchto signálov, čo spolu s fázovým posunom signálov na vstupe vytvorí charakteristiku pásmového priepustu. Tým že jeden fázovací článok je pripojený na vstupy dvoch rozdielových zosilňovačov sú vždy susedné BP filtre fázovo previazané je skreslenie signálov susedných frekvencií výrazne zmenšené.

Rozdielové operačné zosilňovače majú upravené zisky na 15dB. Toto zosilnenie je bežne používané pri profesionálnych ekvalizéroch pre zdvih na jednotlivých frekvenčných pásmach. Za rozdielovými zosilňovačmi sú zaradené obvody digitálnych potenciometrov ktoré regulujú zisk jednotlivých pásiem. Tieto obvody majú vysoký odstup signál/šum, nízke skreslenie a rozdielne napájanie digitálnej a analógovej časti obvodu. Sú ovládané cez I2C rozhranie, umožňujú nastaviť 63 logaritmicky odstupňovaných nastavení. Disponujú aktivačným vstupom zbernice I2C, čo umožňuje multiplexovať väčšie množstvo týchto obvodov pripojených na jednu zbernicu. Adresy týchto obvodov je navyše možné meniť nastavením úrovni na troch riadiacich vstupoch a to v rozmedzí adries 0x50 až 0x5E, čo predstavuje možnosť zapojenia až ôsmych obvodov na jednu zbernicu. Tieto obvody navyše disponujú vnútornou EEPROM pamäťou do ktorej je možné ukladať nastavenia jednotlivých potenciometrov. Za týmito obvodmi nasledujú impedančné prispôsobovače a rezistory, ktoré tvoria vstup súčtového invertujúceho zosilňovača. Tento zosilňovač predstavuje výstup analógovej časti ekvalizéra.

B. Digitálna časť

Jadrom digitálnej časti je moderný procesor STM32F107 [3] s jadrom ARM Cortex-M3, ktorý je v puzdre 64-LQFP. Tento procesor disponuje množstvom rozličných periférií. V aplikácii ekvalizéra je využitá len časť z nich. Hlavnou je I2C zbernica pomocou ktorej sú ovládané digitálne potenciometre. Procesor obsahuje aj USB rozhranie a 3 USART periférie. Jedna z USART periférií je využívaná na nahrávanie firmvéru a konfiguráciu zariadenia. Túto funkciu je možné vykonávať aj cez USB rozhranie, čo bude využité v ďalších verziách firmvéru. Procesor disponuje celkovo až 51 vstupno-výstupnými vývodmi, ktoré využívame na komunikáciu s 2x16 znakovým displejom a ovládacími tlačidlami, pričom každé obsluhuje jeden z vektorov prerušení procesora a jedno z tlačidiel obsluhuje pracovný režim procesora. Na časť vývodov je pripojený prijímač a dekodér infračerveného signálu RC5. V zapojení sú využívané aj PWM výstupy čítačov, ktorými je regulovaná intenzitu podsvietenia displeja a jeho kontrast.

Pre použitie tohto procesora sme sa rozhodli z dôvodu požiadavky väčšieho množstva vstupno-výstupných vývodov a možnosti pripojenia do lokálnej siete, ktoré v budúcnosti umožní konfigurovať navrhnutý ekvalizér aj touto cestou.

Najdôležitejšou súčasťou procesora v aplikácií ekvalizéra je I2C zbernica, ktorá je využívaná na konfiguráciu analógových častí ekvalizéra. Keďže nami použité obvody DS1882 umožňujú zmenu adresy v rozsahu ôsmych zariadení, pre použitý celkový počet BP filtrov akvalizéra by to nebolo postačujúce. V zapojení sú preto využité aj aktivačné vstupy I2C zberníc. Tým sme vytvorili bloky týchto obvodov zložených z ôsmych digitálnych potenciometrov, ktoré sú pripojené na jeden aktivačný vodič. Keďže v celkovom zapojení ekvalizéra potrebujeme 31

potenciometrov na reguláciu ziskov jednotlivých pásiem, jeden potenciometer na reguláciu výstupnej hlasitosti a jeden na reguláciu vstupnej citlivosti, je celkový počet použitých potenciometrov 33. Keďže v jednom puzdre DS1882 sú obsiahnuté dva moduly digitálneho potenciometra, pre ekvalizér potrebujeme 17 obvodov digitálnych potenciometrov. Toto vyžaduje využitie troch aktivačných vodičov. Tieto vodiče sú spoločné pre oba kanály, rozdielne sú iba použité I2C zbernice. Jednou z možností pri návrhu bolo použitie len jednej I2C zbernice a multiplexovanie všetkých obvodov. Touto možnosťou by ale nebolo možné vytvoriť úplne identické moduly analógových častí, čo by zbytočne komplikovalo návrh plošných spojov pre analógovú časť.

C. Zdroj

Napájací zdroj je realizovaný dvoma transformátormi pričom jeden z nich je určený pre analógovú časť ekvalizéra. Tento transformátor je toroidný so symetrickým sekundárnym vinutím a je vypínateľný riadiacim procesorom. Stabilizáciu napätia realizujú obvody LM317 [4] v kladnej vetve napájania a LM337 [5] v zápornej vetve napájania. Pre použitie týchto obvodov sme sa rozhodli z dôvodu generovania nižšieho šumu ako bežne používané lineárne stabilizátory z radu obvodov MA78XX. Druhý transformátor napája digitálne obvody a jeho výstup je stabilizovaný na dve rôzne napätia +3,3V a +5V, keďže použité súčiastky využívajú pre správnu funkčnosť tieto dve napájacie napätia. Stabilizovanie napätia opäť realizujú obvody LM317 v puzdre TO220 [4].

III. POUŽITÉ VÝVOJOVÉ PROSTRIEDKY

Na vývoj riadiaceho programu sme použili vývojové a ladiace prostredie KEIL µVision 4 [6], ktoré obsahuje C/C++ kompilátor, linker a ladiace nástroje. Pri programovaní a ladení riadiaceho programu sme využívali hardvérové ladiace prostriedky Keil uLink [7] a ST-LINK [8]. Pri tvorbe riadiaceho programu sme využili dostupné periférne knižnice od firmy STMicroelectronics, ktoré nám výrazne zjednodušili tvorbu programového vybavenia, keďže pri tvorbe programového vybavenia nebolo potrebné riešiť nízko-úrovňové detaily obsluhy periférnych obvodov procesora. Periférne knižnice sú priebežne aktualizované na stránke výrobcu [9] a demonštračné príklady obsahujú aj aktuálny štartovací súbor.

IV. OPIS ANALÓGOVEJ ČASTI EKVALIZÉRA

Keďže typické zariadenie pre spracovanie audio signálov využíva niekoľko rozličných vstupov kvôli možnosti pripojenia viacerých zariadení ako je PC, TV, DVD a pod., obsahuje navrhnutý ekvalizér elektronický prepínač vstupov. Jeden zo vstupov je symetrický, z dôvodu možnosti pripojenia profesionálnych zariadení, prípadne pripojenia zariadenia na väčšiu vzdialenosť. Prepínač je tvorený trojicou relé ktoré sú budené I2C expandérom. Relé odpájajú aj nulový pól a to z dôvodu zníženia možnosti zanesenia rušenia cez nulové vodiče. Signál pokračuje cez impedančný prispôsobovač s možnosťou nastavenia vstupnej citlivosti a pokračuje do 30 fázovacích článkov. V pôvodnom návrhu [2] bol využitý aj kondenzátor C v spätnej väzbe zobrazený na Obr.2 čiarkovane, ktorý zvýšil prenos v časti frekvenčného pásma o cca 12%. Zároveň však tento kondenzátor mení výstupnú fázovú charakteristiku, čo je pre nás nevýhodné a preto sme tento kondenzátor v návrhu ekvalizéra nepoužili. Na Obr. 2 (vpravo) je možné zreteľne vidieť rozdiel prenosových funkcií dvoch rovnakých fázovacích článkov, pričom jeden je navyše osadený kondenzátorom v spätnej väzbe (červený priebeh). Spodná časť naznačuje fázovo-frekvenčné charakteristiky a je tak isto zrejmé že zapojenie s pripojeným kondenzátorom v spätnej väzbe má rozdielnu aj fázovo-frekvenčnú charakteristiku.

Prenosová funkcia fázovacieho článku na Obr.2 s pridaným kondenzátorom C v spätnej väzbe je určená vzťahom

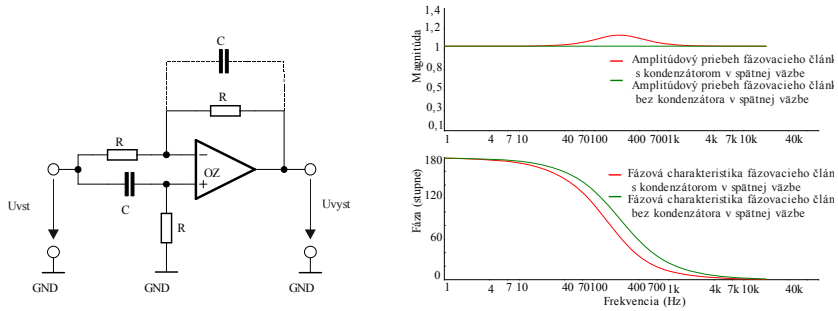
$$A_u = \frac{U_{vyst}}{U_{vst}} = -\frac{1+\omega^2 R^2 C^2 - j\omega RC}{(1+j\omega CR)^2} \tag{1}$$

Pre prípad že frekvencia sa blíži k nule dostávame:

$$\lim_{\omega \to 0} = -1 \tag{2}$$

A pre frekvenciu blížiacu sa k nekonečnu:

$$\lim_{\omega \to \infty} = +1 \tag{3}$$



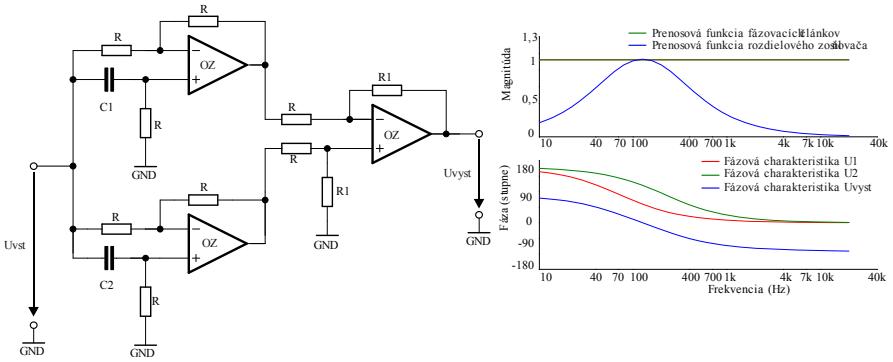
Obr. 2 Základné zapojenie fázovacieho článku(vľavo) a výpočet prenosovej funkcie takéhoto článku(vpravo)

Hraničná frekvencia takéhoto článku je určená vzťahom

$$f_0 = \frac{1}{2\pi RC} \tag{4}$$

BPs filter je tvorený dvoma fázovacími článkami a rozdielovým zosilňovačom. Schéma jedného BP filtra je zobrazená na Obr. 3. Prenosovú funkciu BP filtra je určená vzťahom

$$A_u = \frac{R_1}{R} \frac{j2\omega R(C_2 - C_1)}{(1 + j\omega C_1 R)(1 + j\omega C_2 R)} \tag{5}$$



Obr. 3 Schéma BP filtra s využitím fázovacích článkov (naľavo) a jej priebeh (napravo)

Koncepcia návrhu využíva 1/3 oktávový ekvalizér. V

Tabuľka. 1 sú vypočítané hodnoty všetkých kondenzátorov fázovacích článkov a uvedené aj reálne použité hodnoty kondenzátorov z rady E24, ako aj vypočítané hodnoty vrcholov upravovaných frekvenčných rozsahov.

Tabuľka. 1
hodnoty kapacít kondenzátorov použitých vo fázovacích článkoch, ich hraničných frekvencií, výsledných frekvencií frekvenčných rozsahov a tabuľkové hodnoty 1/3 oktávovej stupnice

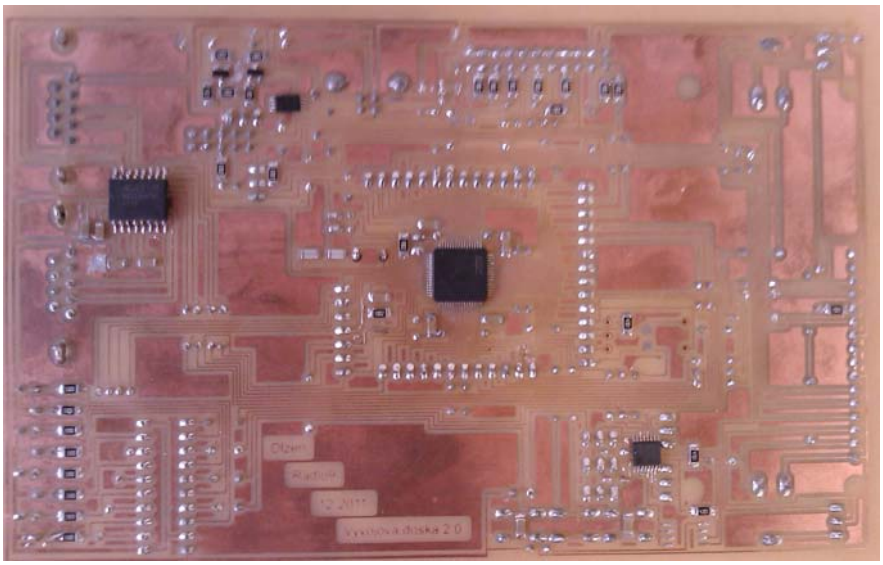
Kapacita kondenzátorov použitých vo fázovacích článkoch	Hraničná frekvencia fázovacích článkov	Výsledná stredná hodnota upravovaného frekvenčného rozsahu	Tretinooktávové frekvenčné pásma
510 nF	14,2 Hz	18,1 Hz	20 Hz
330 nF	21,9 Hz	24,4 Hz	25 Hz
270 nF	26,8 Hz	31,5 Hz	31,5 Hz
200 nF	36,2 Hz	40,7 Hz	40 Hz
160 nF	45,2 Hz	50,4 Hz	50 Hz
130 nF	55,7 Hz	64 Hz	63 Hz
100 nF	72,3 Hz	80,3 Hz	80 Hz
82 nF	88,2 Hz	97,3 Hz	100 Hz
68 nF	106,4 Hz	124,1 Hz	125 Hz
51 nF	141,9 Hz	155,1 Hz	160 Hz
…	…	…	…
0,51 nF	14184,9 Hz	15504,5 Hz	16000 Hz
		19373,1 Hz	20000 Hz

Pri frekvencií 20 kHz ale nejde o pásmový priepust, ale o horno-pásmový priepust, čo znamená že frekvenčné pásmo je upravované od 19,373 kHz po maximum.

Keďže jednou z požiadaviek bola možnosť regulovať zdvih jednotlivých pásiem v rozsahu ±15dB museli sme upraviť prenosovú funkciu rozdielových zosilňovačov. Zo vzťahu (5) sme určili hodnoty rezistorov R1 na hodnotu 3,9125 kΩ pri R=22 kΩ najbližšia hodnota v rade E24 predstavuje 3,9 kΩ.

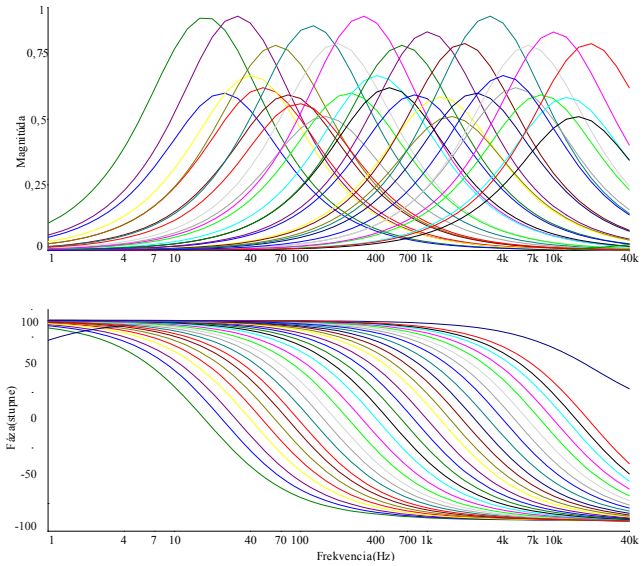
V. EXPERIMENTÁLNE VÝSLEDKY

Navrhovaný ekvalizér a jeho ovládací softvér bol vyvíjaný a testovaný na vyrobenej vývojovej doske s procesorom STM32F107 zobrazenej na Obr. 4. Táto doska obsahovala okrem použitého digitálneho potenciometra DS1882 [10] aj dvoch USART periférií, prevodníkom úrovní, 8 diód spínaných budičom, LCD znakovým displejom a USB rozhraním. Z dôvodu možnosti pripojenia ďalších periférií sú na doske vyvedené všetky vstupno-výstupné vývody procesora.



Obr. 4 Vývojová doska s procesorom STM32F107

Celá schéma analógovej časti bola nasimulovaná v programe Multisim a jej grafický priebeh je zobrazený na Obr. 5.



Obr. 5 Simulácia amplitúdovo-frekvenčných a fázovo-frekvenčných charakteristík BP filtrov ekvalizéra

Grafický priebeh zobrazuje fázovo-frekvenčné charakteristiky (dole) a amplitúdovo-frekvenčné charakteristiky (hore) všetkých BP filtrov. Tým že sú susedné BP filtre fázovo zviazané na nábežnej aj dobežnej časti frekvenčnej charakteristiky BP filtrov, sú fázové charakteristiky použitých BP filtrov výhodnejšie pre použitie v audio technike. Odlišná úroveň prenosu jednotlivých BP filtrov jasne viditeľná na Obr. 5 je dôsledkom použitej topológie BP filtra a výberu R, C prvkov v Tab. 1, ktoré majú vplyv na prenos (5). Táto závislosť musí byť kompenzovaná vhodným nastavením digitálnych potenciometrov v riadiacom softvéri. Aj napriek netradičnej koncepcii navrhnutého ekvalizéra a potrebe kompenzácie nerovnakých ziskov jednotlivých BP filtrov, nie je analógová časť konštrukčne náročná a preto je princíp schémy vhodný pre výrobu v experimentálnych podmienkach.

VI. ZÁVER

V tomto článku sme opísali len základné bloky ekvalizéra ktorého vývoj nie je ešte ukončený. Funkčnosť jednotlivých blokov však bola experimentálne overená pomocou zjednodušeného 3-pásmového ekvalizéra, čo umožnilo overiť navrhnutý princíp riadenia ekvalizéra a doladiť jednotlivé časti riadiaceho softvéru pred finalizáciou kompletnej analógovej dosky finálneho 31-pásmového analógového ekvalizéra. Ďalšie detaily návrhu ekvalizéra sú dostupné v bakalárskej práci [11].

POĎAKOVANIE

Táto publikácia vznikla vďaka podpore v rámci operačného programu Výskum a vývoj, pre projekt: Rozvoj Centra informačných a komunikačných technológií pre znalostné systémy (kód ITMS projektu: 26220120030) na základe podpory operačného programu Výskum a vývoj financovaného z Európskeho fondu regionálneho rozvoja.

REFERENCIE

[1] Wikipedia: Ekvalizér [online]. Dostupné na internete: <http://cs.wikipedia.org/wiki/Ekvaliz%C3%A9r>
[2] J. Míček, „Devítipásmový ekvalizér EKV903“ In: Praktická Elektronika č.10 2003, s22-24.
[3] ST Microelectronics: STM32 Reference Manual [online]. October, 2011, Version 14. Dostupné na internete: <http://www.st.com/internet/com/TECHNICAL_RESOURCES/TECHNICAL_LITERATURE/REFERENCE_MANUAL/CD00171190.pdf>.
[4] ST microelectronics: LM317 Datasheet [online]. Dostupné na internete <http://www.st.com/internet/com/TECHNICAL_RESOURCES/TECHNICAL_LITERATURE/DATASHEET/CD00000455.pdf>
[5] ST microelectronics: LM337 Datasheet [online]. Dostupné na internete <http://www.st.com/internet/com/TECHNICAL_RESOURCES/TECHNICAL_LITERATURE/DATASHEET/CD00000468.pdf>
[6] Keil Product downloads: Keil µVision 4 [online]. Dostupné na internete <https://www.keil.com/demo/eval/arm.htm>
[7] Keil : Ulink debugger and programmer [online]. Dostupné na internete http://www.keil.com/ulink1/
[8] ST Link: ST-Link debugger and programmer [online]. Dostupné na internete <http://www.st.com/internet/evalboard/product/219866.jsp>
[9] ST Link: STM32F10x standard peripheral library [online]. Dostupné na internete <http://www.st.com/internet/com/SOFTWARE_RESOURCES/SW_COMPONENT/FIRMWARE/stm32f10x_stdperiph_lib.zip>
[10] Maxim : Datasheet DS1882-v.0 [online]. June 2006, Dostupné na internete <http://datasheets.maxim-ic.com/en/ds/DS1882.pdf>.
[11] M. Petrinec, „Pásmový ekvalizér s číslicovým riadením“. Bakalárska práca, Košice, TU FEI, 2012.

Sledovanie osôb za stenou pomocou impulzového UWB radaru

Peter PETRUŠKA, Jana ROVŇÁKOVÁ

Katedra elektroniky a multimediálnych telekomunikácií, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská Republika

peter.petruska.2@student.tuke.sk, jana.rovnakova@tuke.sk

***Abstrakt*— Existuje niekoľko typov UWB radarových systémov, ktoré možno použiť za účelom sledovania osôb pohybujúcich sa za prekážkou. Zámerom tohto článku je otestovať v danej aplikácii výkonnosť impulzového UWB radaru v porovnaní s UWB radarovými systémami pracujúcimi na báze M-postupností. Za týmto účelom boli namerané, spracované a analyzované UWB radarové signály získané experimentálnymi cezstenovými meraniami pri použití rozdielnych radarových systémov.**

***Kľúčové slová*—impulzový radar, pohybujúci sa cieľ, sledovanie osôb, spracovanie signálov, UWB radar.**

I. ÚVOD

Zemetrasenie, požiar či lavína si každý rok vyberie svoju daň v podobe vyhasnutých ľudských životov. Sú to faktory, ktorých vznik človek nedokáže vždy ovplyvniť. Ak však k takýmto udalostiam dôjde, je možné následky týchto udalostí zmierniť, a predísť tak najmä stratám na životoch. Jedným zo spôsobov ochrany ľudských životov je použitie UWB radarových systémov [1]. Ich kľúčovou vlastnosťou je dobrá penetračná schopnosť elektromagnetických vĺn vysielaných týmito systémami cez väčšinu typických stavebných materiálov [2]. Použitie radarov pracujúcich vo frekvenčnom pásme do 5 GHz umožňuje detegovať a sledovať pohybujúce sa ciele ukryté za prekážkou (za stenou, pod rozvalinami budov, pod snehom, v hmle, v dyme či v hustom vegetačnom poraste) ako aj lokalizovať žijúce, ale inak nehybné osoby [3]. Radarové systémy tohto typu nachádzajú uplatnenie nielen v oblasti bezpečnosti a ochrany ľudských životov, ale aj pri monitorovaní priestorov pre vojenské a bezpečnostné aplikácie (napr. detekcia a lokalizácia mín pomocou GPR), v stavebníctve, v archeológii pri detekcii artefaktov či dokonca v medicíne pri neinvazívnej diagnostike [2].

UWB radarové systémy pracujú na rôznych princípoch. V poslednom období boli na cezstenové aplikácie s výhodou využívané najmä UWB radary pracujúce na báze M-postupností [3][4][5]. Alternatívou k nim je okrem iného aj impulzový UWB radar [2]. Cieľom tohto článku je otestovať za rovnakých meracích podmienok výkonnosť oboch spomínaných radarových systémov pri sledovaní osôb pohybujúcich sa za stenou [6]. Za týmto účelom je článok rozdelený do troch ďalších kapitol, v ktorých je najskôr definovaná UWB radarová technológia a stručne vysvetlený princíp rôznych radarových systémov. Následne je naznačená metodika spracovania UWB radarových signálov, ktorá je využitá v rámci experimentálnych výsledkov na spracovania radarových signálov nameraných impulzovým radarom ako aj radarom pracujúcim na báze M-postupností. V závere sú diskutované možnosti zlepšenia dosiahnutých výsledkov.

II. UWB RADAROVÉ SYSTÉMY

A. UWB radarová technológia

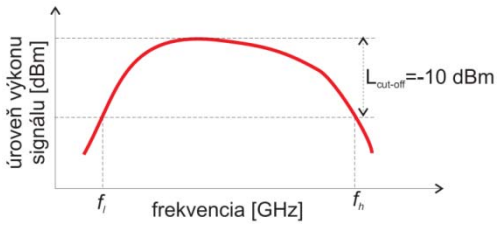
UWB technológie aplikované v radarovej technike umožňujú dosahovať veľmi jemné rozlíšenie radarového systému (napr. radar so šírkou pásma 5 GHz má rozlíšenie cca 3 cm) z čoho vyplýva jeho potenciálna schopnosť veľmi presne lokalizovať ciele, možnosť detegovať mnohonásobné ciele, ako aj vyššia odolnosť radarového systému proti rušeniu.

Ak je zlomková šírka frekvenčného pásma signálov emitovaných radarom väčšia ako 0.2 GHz, alebo ak je ich absolútna šírka frekvenčného pásma väčšia ako 0.5 GHz, hovoríme o radaroch s veľmi veľkým rozlíšením (z angl. high-resolution radars) alebo o radaroch

pracujúcich s veľmi veľkou šírkou frekvenčného pásma (z angl. UltraWideBand, UWB)[1][2]. Absolútna šírka pásma B a zlomková šírka pásma B_f sú definované vzťahmi

$$B = f_u - f_l \,, \quad B_f = \frac{2(f_u - f_l)}{f_u + f_l} \,,$$

kde f_u a f_l sú horná a dolná hraničná frekvencia určená hraničnou úrovňou signálu $L_{cut\text{-}off} = -10\ dBm$. Na Obr.1 je znázornený význam parametrov použitých na určenie zlomkovej a absolútnej šírky frekvenčného pásma [4].



Obr.1 Parametre určujúce absolútnu a zlomkovú šírku frekvenčného pásma signálu [4].

B. Radarové systémy

Existuje niekoľko druhov radarových systémov:

- Impulzový UWB radar
- Impulzový Dopplerov radar
- Frequency Modulated Continuous Wave (FM-CW) radar
- Radar so stupňovitou frekvenčnou moduláciou (Stepped-FM) radar
- Šumový radar
- Radar na báze M-postupností

Impulzový UWB radar pracuje tak, že veľmi krátky impulz (kratší ako 1ns) je privádzaný priamo na vysielaciu anténu, ktorá sa správa ako pásmový priepust. Tvar a šírka pásma vyžarovaného signálu sú určené prenosovou funkciou antény, t.j. časťou spektra, ktorú môže anténa efektívne prenášať. Pre tvarovanie spektra sa môže pred anténu zaradzovať filter. Čím je impulz kratší, tým je jeho spektrum širšie a so skracujúcou sa dobou trvania na nulu sa jeho spektrum blíži k spektru bieleho šumu. Prijímaný signál je potrebné vzorkovať a hodnoty uchovávať pre ďalšie spracovanie.

Impulzový Dopplerov radar vysiela sériu krátkych impulzov, ktoré sa čiastočne odrazia od pozorovaných objektov. Zmeraním Dopplerovho posunu signálu odrazeného od objektu je radar schopný extrahovať informáciu, napríklad o rýchlosti pohybu objektu. Použitie krátkych impulzov namiesto stálej vlny znižuje výkonové nároky a výpočtovú zložitosť. Príkladom využitia je policajný radar merajúci rýchlosť vozidiel.

FM-CW radar je radar s kontinuálne frekvenčne modulovanou vlnou. Mení frekvenciu vysielaného signálu a meria vzdialenosť na základe rozdielu frekvencií medzi práve vysielaným signálom a prijímaným odrazeným signálom. Modulačnými signálmi sú typicky signály s pílovým alebo trojuholníkovým priebehom [2].

Stepped-FM radar prijíma echo od cieľa, ktoré prechádza fázovo citlivým synchrónnym detektorom (tiež označovaný I/Q demodulátor, kvadratúrny detektor, koherentný detektor), čím sa získa rad úzkopásmových impulzov v základnom pásme [2].

Šumový radar funguje na princípe vysielania signálu náhodného šumu, ktorý je od objektu odrazený späť a vstupuje do prijímača. Prijatý signál je zosilnený a korelovaný s odoslaným signálom [2].

Radar na báze M-postupností periodicky vysiela nosný signál pseudonáhodnej postupnosti s maximálnou dĺžkou.

Podrobnejší opis fungovania jednotlivých radarových systémov je možné nájsť v [2][7][8][9].

III. METODIKA SPRACOVANIA UWB RADAROVÝCH SIGNÁLOV

Zvolená metodika bola pôvodne vytvorená pre spracovanie signálov získaných UWB radarom pracujúcim na báze M-postupností. Prvotné výsledky spracovania signálov nameraných impulzovým UWB radarom však potvrdili, že ju možno s výhodou využiť aj na spracovanie takýchto radarových signálov.

Spracovanie nameraných radarových signálov prebiehalo v piatich etapách:

- Odčítanie pozadia
- Detekcia cieľa
- Odhad času príchodu signálu (Time-of-arrival, TOA)
- Lokalizácia cieľa
- Sledovanie pohybu cieľa

Odčítanie pozadia zlepšuje odstup užitočného signálu od šumu a vytvára tak lepšie podmienky pre detekciu cieľa.

Detekcia cieľa je etapa, ktorá má určiť, či možno v signáli po odčítaní pozadia identifikovať potenciálny cieľ na základe zvolených testovacích hypotéz.

Odhad času príchodu signálu má za úlohu nájsť prvý odraz od každého cieľa a vytvoriť skupinu takých odrazov cez všetky prijímacie antény, ktoré prislúchajú tomu istému cieľu.

Lokalizácia cieľa určuje po odhade času príchodu signálu pozíciu cieľa v zvolenej súradnicovej sústave.

Sledovanie pohybu cieľa sa využíva v prípade pohybujúceho sa cieľa, kde na základe informácií o polohe cieľa v aktuálnom časovom okamihu a predchádzajúcich časových okamihoch je možné dosiahnuť spresnenú trajektóriu pohybu cieľa.

Princíp a použitie jednotlivých etáp je podrobne opísaný v [3][4].

IV. Experimentálne výsledky

A. Meracia kampaň

Za účelom získania databáz radarových signálov zo všetkých troch radarových systémov dostupných na KEMT, TUKE bola v priestoroch katedry uskutočnená rozsiahla meracia kampaň. Pre všetky radarové systémy, t.j. impulzový radarový systém a radarové systémy na báze M-postupností s dosahom 17 m a 47 m, boli zabezpečené rovnaké podmienky merania. Každý použitý radarový systém mal jeden vysielací (Tx) a dva prijímacie kanály (Rx1 a Rx2). Vo všetkých prípadoch boli použité otvorené antény typu Horn (Obr.2).



(a) (b) (c)

Obr.2 Použité UWB radarové systémy: (a) radarový systém na báze M-postupností s dosahom 17 m, (b) radarový systém na báze M-postupností s dosahom 47 m, (c) impulzový radarový systém.

Prvým použitým systémom bol radar na báze M-postupností s dosahom 17 m. Využíva širokopásmové signály so vzorkovacou frekvenciou 4.5 GHz a šírkou frekvenčného pásma 0 ~ 2.25 GHz. Pri tomto type radaru je rýchlosť merania približne 14 impulzových odpovedí (z angl. Impulse Response, IR) za sekundu. Perióda vzorkovania bola nastavená na 511 vzoriek (Obr.2(a)).

Druhý radar s dosahom 47 m bol rovnako systémom na báze M-postupností využívajúci ultraširokopásmové signály so vzorkovacou frekvenciou 13 GHz a šírkou pásma 0.2 ~ 6.5 GHz. Rýchlosť merania je 84.5 IR za sekundu a perióda vzorkovania môže dosiahnuť maximálnu hodnotu 4095 vzoriek. Pre tieto merania sme využili všetkých 4095 vzoriek, ale pri spracovaní sme skrátili maticu dát na 511 vzoriek, pretože cieľ sa nachádzal v dostatočnej blízkosti radaru (Obr.2(b)).

Impulzový UWB radarový systém bol použitý v režime dvoch prijímacích kanálov Rx1 a

Rx2 a počas meraní sme jeho dosah regulovali v rozsahu približne 5 až 17 m. Vzorkovacia frekvencia bola nastavená na 500 kHz. Hodnoty parametrov (napr. vzorkovacia frekvencia, počet použitých vzoriek v časovom okne, atď.) boli určené na základe predchádzajúcich experimentálnych meraní. Použité časové okno bolo 100 ns a oneskorenie 44 ns. Počet použitých vzoriek v časovom okne bol 512, čím sa dosiahlo zvýšenie počtu zachytených IR v porovnaní s väčším počtom vzoriek ako 512 v časovom okne. Impulzový UWB radar umožňuje aj použitie externých zosilňovačov vstupných kanálov, avšak cieľom tejto práce bolo porovnať vlastnosti daného radaru bez zapojenia akýchkoľvek externých zosilňovacích prvkov (Obr.2(c)) [6].

Za sledované priestory (monitorovanú oblasť) boli zvolené dve susedné miestnosti, a to takmer prázdna trieda s rozmermi 7.35 m x 5.35 m (Obr.3(a)) a jedáleň s dĺžkou 21 m plne zariadená stolmi a stoličkami (Obr.3(b)).

Radarové systémy spolu s anténnym systémom boli umiestnené v triede, v ľavom zadnom rohu miestnosti (Obr.3(a)). V prednej časti triedy sa nachádzala tabuľa (pôsobiaca ako silný reflektor) s rozmermi 2 m x 1.2 m. Hrúbka steny medzi jedálňou a triedou bola 35 cm.



(a) (b)

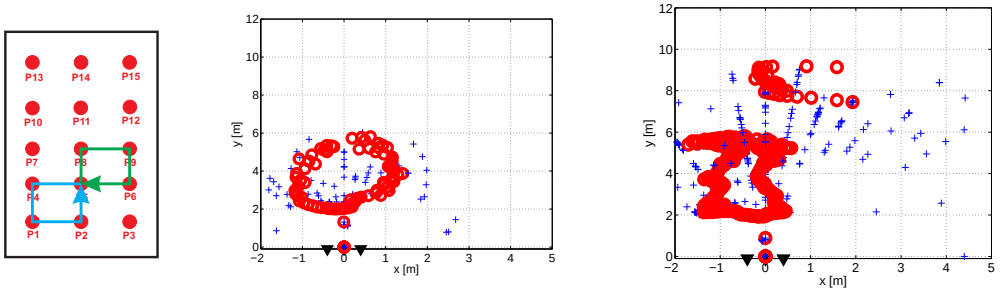
Obr.3 Fotografie sledovaných priestorov: (a) trieda, (b) jedáleň.

B. Scenáre

Z celkového súboru nameraných signálov bol vybraný ako príklad scenár s označením *f*, ktorý demonštruje výkonnosť UWB impulzového radaru pri detekcii viacerých pohybujúcich sa cieľov pri cezstenovom meraní v plne zariadenej miestnosti. Išlo o scenár, v ktorom dve osoby kráčali v jedálni po oddelených obdĺžnikových trajektóriách v rozdielnych vzdialenostiach od radarového systému nachádzajúceho sa za stenou (Obr.4(a), Obr.3(b)). Dosiahnuté výsledky sú opísané a analyzované v nasledujúcich sekciách.

C. Výsledky

Na Obr.4 sú porovnané výsledky spracovania radarových signálov nameraných impulzovým radarom a radarom na báze M-postupností. Oba radary pracujúce na báze M-postupností poskytovali pre tento scenár veľmi podobné výsledky, preto bol v rámci porovnania s impulzovým radarom vybraný iba radar s dosahom 17 m (MR17). Z Obr.4(b) a Obr.4(c) je zrejmé, že oba radary boli schopné zachytiť úplnú trajektóriu iba bližšieho cieľa. Odrazy od vzdialenejšieho cieľa boli iba útržkovite zachytené radarom MR17 (Obr.4(c)), pričom impulzový radar nedetegoval druhý cieľ vôbec (Obr.4(b)). Dosiahnuté výsledky vyplývajú z charakteru nameraných a spracovaných radarových signálov (Obr.5 a Obr.6).



(a) (b) (c)

Obr.4 Trajektórie scenára f: (a) skutočná, (b) odhadnutá impulzovým radarom, (c) odhadnutá radarom MR17.

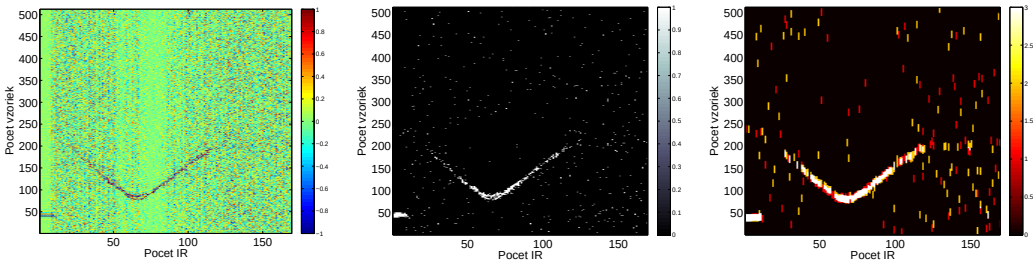
D. Analýza

Na Obr.5 a Obr.6 sú znázornené radargramy signálov z etapy odčítania pozadia, detekcie cieľa a odhadu TOA. Ide o dvojrozmerné zobrazenia radarových signálov, v ktorých horizontálna os korešponduje s počtom nameraných impulzových odpovedí a vertikálna os zodpovedá počtu vzoriek jednej impulzovej odpovede. Farebnou škálou je vyjadrená hodnota spracovaných dát, ktorej význam sa v jednotlivých etapách mení.

Radargramy po odčítaní pozadia nadobúdajú po normovaní hodnoty od -1 po 1, pričom okrajové hodnoty tejto stupnice zodpovedajú najsilnejším odrazom od cieľov. Porovnaním Obr.5(a) a Obr.6(a) je možné pozorovať, že impulzový radar prijal podstatne zašumenejšie signály, čo malo za následok prítomnosť viacerých falošných odrazov po etape detekcie cieľa (Obr.5(b)).

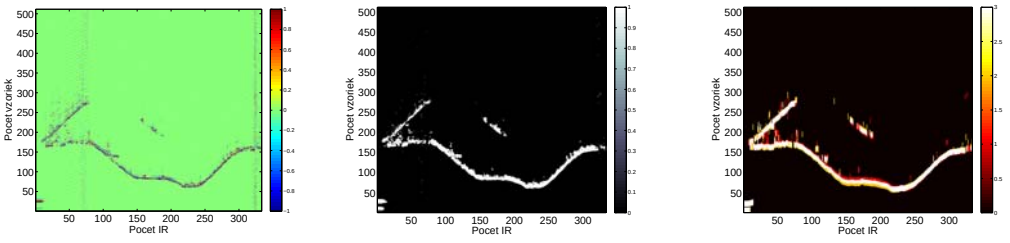
Odpoveď detektora nadobúda vo všeobecnosti binárne hodnoty, v ktorých hodnoty 1 vyjadrujú prítomnosť cieľa a prepočtom vzhľadom na vertikálnu os radargramu je možné vypočítať vzdialenosť príslušného cieľa od anténneho systému radaru. Odpoveď detektora pre radar MR17 (Obr.6(b)) je bohatšia v porovnaní s impulzovým radarom (Obr.6(a)), lebo naznačuje manévrovanie bližšieho cieľa ako aj prítomnosť druhej osoby.

Radargramy s odhadnutým TOA nadobúdajú hodnoty 0, 1, 2 alebo 3, pričom ich význam je nasledovný: 0 – neprítomnosť cieľa (čierna farba v radargrame na Obr.5(c) a Obr.6(c)), 1 – stopy cieľov odhadnuté z Rx1 (červená farba), 2 – stopy cieľov odhadnuté z Rx2 (žltá farba), 3 - stopy cieľov potvrdené obidvoma kanálmi (biela farba). Pod stopou cieľa tu rozumieme umelo rozšírené prvé odrazy od cieľov, ktoré sú získané v etape odhadu TOA. Na Obr.5(c) môžeme pozorovať, že v prípade využitia impulzového radaru ani stopa bližšieho cieľa nevytvárala spojitú krivku, čo malo za následok chýbajúce pozície cieľa počas etapy lokalizácie. Radar MR17, Obr.6(c), odhadol stopu bližšieho cieľa takmer ukážkovo, čo umožnilo sledovať cieľ aj pri ostrých zmenách smeru, druhý cieľ bol zachytený aspoň čiastočne.



(a) (b) (c)

Obr.5 Radargramy scenára f pre impulzový radar: (a) odčítanie pozadia z Rx1, (b) detekcia cieľa z Rx1, (c) odhad TOA z oboch Rx (Rx1-červená farba, Rx2-žltá farba).



(a) (b) (c)

Obr.6 Radargramy scenára f pre radar MR17: (a) odčítanie pozadia z Rx1, (b) detekcia cieľa z Rx1, (c) odhad TOA z oboch Rx (Rx1-červená farba, Rx2-žltá farba).

V. ZÁVER

Analýza spracovaných UWB radarových signálov z obdobných scenárov, ako bol demonštrovaný v experimentálnych výsledkoch, ukázala, že výsledky dosiahnuté impulzovým radarom nie sú vo väčšine prípadov tak dobré ako u UWB radarov pracujúcich na báze M-postupností. Impulzový radar dosahoval najlepšie výsledky pri sledovaní cieľa do vzdialenosti približne 5 metrov od prekážky. Ciele vzdialené za touto hranicou boli detegované chybne, prípadne neboli detegované vôbec. Tento rozdiel v porovnaní s radarmi pracujúcimi na báze M-postupností je pravdepodobne spôsobený malým SNR, čo by mohlo zlepšiť použitie externých zosilňovačov s vyšším ziskom a nižším šumom. Zlepšenie výsledkov je možné dosiahnuť aj lepším softvérovým doladením parametrov systému, ktoré by umožnilo získať napr. viac impulzových odpovedí za sekundu.

POĎAKOVANIE

Táto práca bola podporovaná Agentúrou na podporu výskumu a vývoja na základe zmluvy č. LPP-0080-09 a Kultúrnou a edukačnou grantovou agentúrou Ministerstva školstva, vedy, výskumu a športu Slovenskej republiky na základe projektu č. 010TUKE-4/2012.

REFERENCIE

[1] TAYLOR, J. D.: Ultra wideband Radar Technology. CRC Press, 2000.
[2] KOCUR, Dušan et al.: UWB bezdrôtové senzorové siete. [online]. Fakulta elektrotechniky a informatiky TU v Košiciach, December 2011. [cit. 2012-04-04]. Dostupné na nternete: <https://moodle.tuke.sk/moodle202/mod/resource/view.php?id=1657>.
[3] ROVŇÁKOVÁ, Jana: Spracovanie signálov umožňujúce sledovanie cieľov pohybujúcich sa za stenou prostredníctvom UWB radarového systému s M-postupnosťou. Dizertačná práca. Fakulta elektrotechniky a informatiky TU v Košiciach, august 2009.
[4] ŠVECOVÁ, Mária: Lokalizácia cieľa UWB radarovým systémom. Dizertačná práca. Fakulta elektrotechniky a informatiky TU v Košiciach, august 2009.
[5] AFTANAS, Michal: Zobrazovanie objektov cez stenu pomocou UWB radarového systému. Dizertačná práca. Fakulta elektrotechniky a informatiky TU v Košiciach, august 2009.
[6] PETRUŠKA, Peter: Metodika merania a spracovania UWB radarových signálov. Diplomová práca. Fakulta elektrotechniky a informatiky TU v Košiciach, máj 2012.
[7] KOCUR, Dušan: Ročná správa o riešení projektu - 2007. Fakulta elektrotechnikya informatiky Technickej univerzity v Košiciach, december 2007.
[8] SEKERÁK, Marek: Pozičné senzory pre skenovanie objektov cez steny pomocou UWB radarových systémov. Bakalárska práca. Fakulta elektrotechniky a informatiky TU v Košiciach, 2009.
[9] IMMOREEV, I., "Signal Processing in Ultra Wide Band (UWB) Radars," Fifth International Conference on Radar Systems - RADAR '99, France, Brest, 1999.

ECC Cryptographic Library for ARM Processors

[1]*Martin PETRVALSKÝ,* [2]*Miloš DRUTAROVSKÝ*

Department of Electronics and Multimedia Communications, Faculty of Electrical Engineering and Informatics Technical University of Košice, Slovak Republic

[1]martin.petrvalsky@student.tuke.sk, [2]milos.drutarovsky@tuke.sk

Abstract **– In this paper we present developed cryptographic library based on elliptic curves and optimized for selected cores of embedded ARM processors. We briefly describe used hardware, software and theoretical mathematical background required for this type of cryptography. Our library consists of the low-level functions for addition, substraction and multiplication of large numbers. Routines for Galois field arithmetic and elliptic curve points operations form the middle level of the library. On the top of the library we included protocols for encryption and decryption. We used C language as much as possible, only for a few time-critical low-level routines we used hand-optimized assembler. Our critical library routines include also countermeasures against several well known side-channel attacks. We also compare the performance of implemented library with selected hird party solutions.**

Keywords **– ARM, ECC, library, processors, C-compiler, hand-optimized assembler**

I. INTRODUCTION

The advantage Elliptic Curve Cryptography (ECC) compared to other public key cryptosystems is that it uses relative small sizes of keys. For example if we take RSA (Rivest Shamir Adleman) algorithm we need five times longer keys (see Table 1.1 in [1]). The most common and also the most critical operation in cryptographic systems is multiplication. Smaller keys means faster operations. In conclusion when we consider equal security level, speed and power consumption ECC is the best option among current public key algorithms.

We choose the low-end processors from ARM for our library. They are widely used in embedded applications and they have low power consumption. We use older ARM7TDMI in order to compare our implementation results with [2]. Then we select modern ARM Cortex-M0, ARM Cortex-M3 processors and ARM Cortex-M1 soft-core one optimized for FPGA circuits as very perspective ones for modern embedded designs.

Nowadays we can find some ECC libraries that are especially designed for ARM processors. For example they are *INVIA - ECC Software Library IP* [3] or *Cryptographic library for ARM7TDMI processors* by Jaroslav Bán [2]. These libraries often provide very high speed optimization. For example almost every part of [2] is programmed in assembler. This is a disadvantage when we need to port library to another target.

Our goal is to create ARM ECC library (AEL) and find an optimum among portability, speed and security of the implementation. We take approach from the bottom to the top so we develop low-level function first and the highest level functions at the end. We create implementation in C code first in order to have some reference implementation. With this reference implementation we can later compare optimized parts of our library.

We choose to implement only NIST ECC curves [1]. These curves have speed advantage over general ones as we can use fast algorithms for them. We also choose only one Galois field. $GF(\mathbf{p})$ (Galois Field over prime numbers) suits our processors the best when we consider speed and code sizes. Our goal is also to minimize the memory requirements so we implement only general ECC functions. Indeed we could also use some window functions or unpacked algorithm but it is not our goal in this stage of the AEL development.

Our next objective is optimization in terms of speed and security while provide improved portability. Firstly we need to find bottlenecks of our implementation. For this purpose we can use performance analyzer included in the used software tools. Then we use hand-optimized assembler code to speed up a small set of critical low-level library routines. After speed optimization we take a look at security optimization. We need to add effective countermeasures to higher levels of our library in order to secure it against known side-channel attacks.

The last goal is to test the library. We make sure that all routines give correct result. We check the speed, code sizes of all algorithms and we compare our implementations to the selected ARM based ECC libraries.

II. Hardware and Software

A. Target Microcontrollers

First selected processor is ARM7TDMI [4]. It uses ARM architecture version 4 (ARMv4 architecture is now obsolete). Abbreviation TDMI stands for:

- Thumb - 16-bit Thumb instruction set
- Debugging - On-chip debugger
- Multiplier - 64-bit result powerful multiplier
- ICE - EmbeddedICE hardware

We work with MCBSTM2130 evaluation board by Keil. It contains Phillips (now NXP) LPC2138 microcontroller (MCU) which is based on ARM7TDMI-S core.

The Cortex-M0 [5] processor is designed especially for low power consumption and it has low level gate count. This is required for deeply embedded applications for which this MCU suits best. Compared to ARM7TDMI core this processor implements ARMv6 architecture. M0 is multistage, configurable, 32-bit RISC processor. It executes Thumb/Thumb-2 code subset and it is compatible with other Cortex-M cores. ARM Cortex M0 also has:

- Very tight integration of peripherals in order to minimize area and development costs
- Power control optimization of system components
- Sleep mode functions for power savings
- Configurable hardware multiplier with 32-bit result
- Serial wire debug
- Deterministic and high-performance interrupt handling for time-critical applications

We test Cortex-M0 core on the evaluation board LPCXPresso1113 which is based on NXP LPC1113 MCU.

Cortex-M1 [6] is the first ARM processor especially designed for FPGA implementations. It can be ported to all major FPGA targets and includes support for leading FPGA synthesis tools. It uses ARMv6-M Thumb instruction set with some of Thumb2 instructions, similar to Cortex-M0 processor. ARM Cortex-M1 is compatible with FPGA tools:

- Actel Libero
- Altera Quartus-II
- Synopsis Synplify Pro
- Mentor Precision
- Xilinx ISE

The last processor on our list is ARM Cortex-M3 [7] with ARMv7-M architecture. It is low-power processor with low interrupt latency and low-cost debug useful mainly for deeply embedded applications where latency of interrupts are critical. More features of Cortex-M3:

- Thumb and subset of Thumb-2 instruction set
- Hardware divide instructions
- Banked stack pointer only
- Handler and Thread mode, Thumb and Debug states
- Memory protection unit
- JTAG (Joint Test Action Group) and SWD (Serial Wire Debug) for low-cost debugging

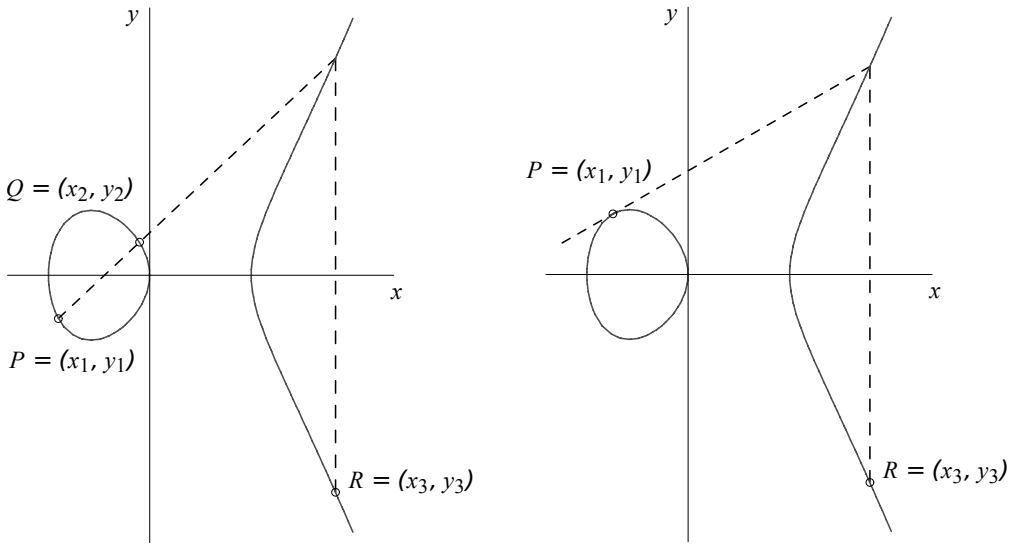
We test Cortex-M3 core on the Keil's MCBSTM32 evaluation board equipped with STM32F103 MCU.

B. Software Development tools

We use development environment μVision by Keil [8] as a part of the MDK-ARM package. For compiling the library and test routines we use IDE Keil μVision version 4.50 with ARM CC compiler V4.1.0.894. This environment combines four features which are needed or helpful for efficient program development:

- Project management and editing
- Flexible compilation options
- Source code editor
- Environment for debugging and program simulation

We also use CodeSourcery Source CodeBench Lite for ARM GCC (GNU C Compiler) by Mentor Graphics [9]. It is specialized for ARM processors and theirs architectures and it is also supported by MDK-ARM as an external compiler. We use Sourcery CodeBench Lite v. 2011.09-69 for compilation of all low-level routines and optionaly also for compilation of testing projects.



(a) Addition: $P + Q = R$. (b) Doubling: $P + P = R$.

Fig. 1 Geometrical interpretation of ECC point (a) adding, (b) point doubling.

III. Elliptic Curve Cryptography

Elliptic curves in mathematics are defined by Weierstrass equation [1]. In ECC we need only the simplified version of this equation

$$\mathbf{EC} : y^2 = x^3 + ax^2 + b \tag{1}$$

where $a, b \in K$ and $\Delta = -16(4a^3 + 27b^2)$. Δ is the discriminant of the curve and it must be a non zero number. All operations such as addition or multiplication have to be performed in a finite field arithmetics. For our purpose we chose Galois field over prime numbers denoted as $GF(\mathbf{p})$.We must define one more point on the curves. It is point at infinity (denoted by ∞) which is the only point on the line at infinity that satisfies elliptic curve equation.

In this work we use only NIST curves [10] specially chosen by NIST organization. There are 5 recommended prime fields determined by 5 prime numbers:

- $\mathbf{p}_{192} = 2^{192} - 2^{64} - 1$
- $\mathbf{p}_{224} = 2^{224} - 2^{96} + 1$
- $\mathbf{p}_{256} = 2^{256} - 2^{224} + 2^{192} + 2^{96} - 1$
- $\mathbf{p}_{384} = 2^{384} - 2^{128} - 2^{96} + 2^{32} - 1$
- $\mathbf{p}_{521} = 2^{521} - 1$

Points on elliptic curve are points with coordinates which satisfies (1). Basic point coordinates (x, y) are called *affine*. There are also another types of coordinate systems which can improve speed or memory requirements of ECC. These are for example *Jacobian*, *Co-Z* or *OnlyX* coordinate systems. We can perform addition, subtraction, doubling and most important multiplication with these points. Addition and subtraction can be easily explained geometrically as shown on Fig. 1. We can imagine point multiplication $k * \mathbf{P}$ (where k is an integer and $\mathbf{P}$ is a point on elliptic curve) as multiple adding of $\mathbf{P}$

$$k * \mathbf{P} = \mathbf{P} + \mathbf{P} + \mathbf{P} + ... + \mathbf{P} \tag{2}$$

where number of added $\mathbf{P}$ points are exactly k. In real applications we should use another algorithms such as left-to-right, right-to-left binary method or Montgomery ladder for efficient evaluation of (2).

Point multiplications are used by ECC protocols which provide encryption, decryption, digital signature or key establishment. All implemented protocols using ECC are based on the Elliptic Curve Discrete Logarithm Problem (ECDLP).

ECDLP is: given an elliptic curve $\mathbf{EC}$ *defined over* $GF(\mathbf{p})$*, a point* $\mathbf{P}$ *on curve* $\mathbf{EC}$ *of order* n *and point* $\mathbf{Q} \in \{\mathbf{P}\}$ *(set generated by* $\mathbf{P}$*), find the integer* $l \in [0, n-1]$ *such that* $\mathbf{Q} = l * \mathbf{P}$*. The integer* l *is called discrete logarithm of* $\mathbf{Q}$ *to the base* $\mathbf{P}$*, denoted* $l = log_{\mathbf{P}}\mathbf{Q}$*.*

In this work we use only ECIES (Elliptic Curve Integrated Encryption Scheme) protocol which can encrypt and decrypt messages. We use this protocol only as an example how elements of AEL can be efficiently used.

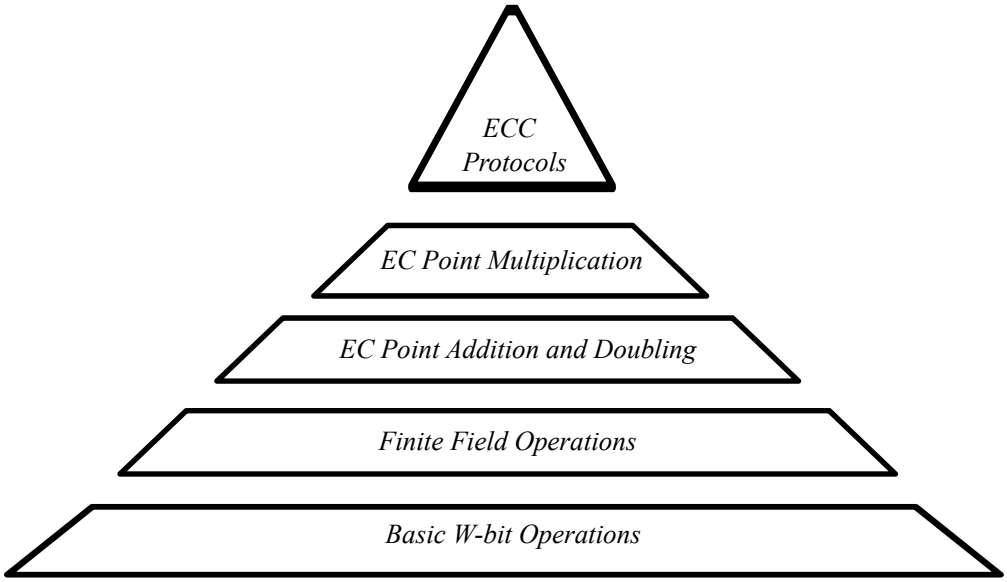


Fig. 2 Elliptic curve cryptography pyramidal hierarchy.

IV. AEL Library

We use approach of creating library from the bottom to the top. First we create whole library in C language - basic multi-precision routines, $GF(\mathbf{p})$ arithmetics, elliptic point operations and the last one is the ECIES protocol to demonstrate AEL library usage.

After having entire library written in C we can continue with next goal which is optimization of the low-level routines in assembler language. Then we focus on securing library against side-channel attacks. The heart of ECC is point multiplication and it is very vulnerable to these kind of attacks. We use some algorithms and techniques which minimize side-channel leaks and can also improve other security issues.

The end of this section belongs to the AEL library testing. We design test programs to prove that our results are correct. Also we measure algorithm speed and we compare results with other implementations.

A. C code library implementation

We start with the lowest-level routines as shown on Fig. 2 and then we continue to higher levels of the pyramid. We have to develop these functions in C code:

- Basic operations - multiprecision addition, subtraction, multiplication, halving, compare, move and pseudo-random number generator functions
- $GF(\mathbf{p})$ operations - addition, subtraction, multiplication, inversion and reduction functions
- ECC point operations - addition, doubling and multiplication
- ECC protocols - ECIES protocol as an example to show how ECC multiplying can be used in higher routines and it uses XTEA (eXtended Tiny Encryption Algorithm) primitives for symmetric encryption, decryption, MAC and key generation

When we have working library written in C code we can use it as a reference for next steps which are optimization and library testing. Also we port this reference library to the PC platform in order to test the library on different CPU architecture and make sure that our implementation is fully portable.

B. Optimization of library

We use two approaches for optimization. First one is low-level routines speed up. We have to analyze where bottlenecks of the library are. For this purpose we use performance analyzer which is part of used μVision IDE. The most critical routines in terms of speed are multiprecision multiplication, addition and subtraction. For each ARM MCU core we have to analyze the instruction set and then optimize time-critical routines using hand-optimized assembler code.

We need to use only additions with carry or subtractions with borrow when we are optimizing multiprecision addition and subtraction. For multiplication we can chose different ways to reach our goal [11]. We use two methods in our library. They are hybrid and Comba methods. Assembler

Table 1
Basic information about ARM ECC Library performance

CPU	Code size	ECIES encryption (256 bit keys)	ECIES decryption (256 bit keys)
ARM7TDMI	12,690B	26,921,818 cycles	14,302,535 cycles
Cortex-M3	7,948B	23,900,818 cycles	12,695,380 cycles
Cortex-M0/M1	8,360B	43,920,133 cycles	23,333,680 cycles

code is about 50% faster then C code for ARM7TDMI and Cortex-M3. For Cortex-M0 and Cortex-M1 it is about 25% faster. These implementations are very close to the architecture maximum performance.

The second approach in optimization is focused on the security. There are four main types of attacks. Protection against brute-force and mathematical attacks can be done by choosing sufficiently large prime number. Power analysis and safe error attacks can be eliminated by right chosen algorithms. The most vulnerable routine for side channels leaks is point multiplication algorithm. We take approach as it is presented in [12]. We use Montgomery ladder and point randomization to countermeasure DPA and SPA (Differential and Simple Power Analysis). In order to speed up this implementation and save memory requirements we use special coordinate system which uses only x-coordinate and (3) for joined point add and double operations.

$$\begin{aligned} T'_D &= T_D W \\ T'_a &= T_a W^2 \\ T'_b &= T_b W^3 \\ X'_1 &= V[(X_1 + X_2)(X_1^2 + X_2^2 - U + 2T_a) + T_b] - T'_D \\ X'_2 &= U[(X_2^2 - T_a)^2 - 2X_2 T_b] \end{aligned} \tag{3}$$

where

$$\begin{aligned} T_a = aZ^2; T_b = 4bZ^3; T_D = x_D Z; U = (X_1 - X_2)^2; \\ V = 4X_2(X_2^2 + T_a) + T_b; W = UV \end{aligned}$$

C. Library testing

We created four functions to perform library testing:

- *test_MUL()* - tests all multiplication routines.
- *test_INV()* - functions for inversion routines tests.
- *test_ECC()* - testing ECC point operations.
- *test_PRO()* - performs tests of ECIES protocol.

All these functions are using timers built in target MCUs to measure number of cycles required to execute analysed operations. They also write log and measured cycles log reports to the serial port of targe MCU so we can watch, save and evaluate the results. To evaluate ECC point operation tests we use also Magma Calculator which can be found at http://magma.maths.usyd.edu.au/calc/.

Finally we evaluate measured results of library testing. There are several aspects to compare (multiprecision multiplication can be a good reference to do these comparisons):

1) CPU comparison - Cortex-M3 and ARM7TDMI processors have very good speed. Cortex-M0 and Cortex-M1 processors need three times more cycles but code is much smaller.
2) Simulator, Flash and RAM comparison - Simulator is always the fastest because it does not count actual bus wait states. Running code from RAM is the second best and the slowest is the execution from Flash. The differences between execution times for targets are very small for Cortex-M3 and ARM7TDMI but it plays big role for Cortex-M0 and M1.
3) C and assembler codes comparison - Optimization with assembler can save number of cycles and even code size. For ARM7TDMI and Cortex-M3 we managed to save approximately 50% of execution time and 25% for Cortex-M0 and M1.

Basic information about AEL performance can be found in Table 1. RAM utilization for ECIES encryption depends on used curve and some other parameters. It can be estimated as

$$ECIESRAM = LEN * 17 + LENM/2 + KEY * 2 + MAC * 2 + 7 \tag{4}$$

where LEN is the number of 32 bit words used to store one multi-precision number, LENM is the length of a message in chars, KEY and MAC are length of key and length of MAC code generated by KDF function. Maximum RAM used is for 521 bit length words and it is *2.5 kB*.

To compare results of our library with ACL library [2] we need speed of ECC point multiplication on ARM7TDMI. Using 256 bit long words the multiplication takes 13,744,840 cycles and with security optimization it is 15,033,796 cycles. It is more then two times slower than ACL implementation but it is the price to pay for much better portability. Almost all codes from ACL point multiplication was programmed in assembler. In terms of code sizes both libraries are comparable. If we subtract code sizes of curves that were used in ACL we get almost the same result.

Comparison with INVIA library [3] can be also interesting. The results they offer are for ARM Cortex-M3 core. Encryption (using ECIES and 256 bit long words) in 86 ms on ARM Cortex-M3 at 100 MHz could be recounted as 8,600,000 cycles. Compared to our implementation it is 3 times faster but as you can see the code size and RAM used are at least 2 times larger. We can assume that INVIA implementation is using methods to speed up which take a lot of space in RAM and in Flash. For example they can use window functions or pre-computed values stored in memories which make code much faster but it also takes a lot of Flash and RAM space.

V. Conclusion

We have created elliptic curve cryptography library for various types of MCUs like ARM7TDMI, ARM Cortex-M3, M1 and M0. It supports all NIST curves over primary field arithmetics. Our library is optimized for speed, security and portability. Almost whole library is written in C language. Only a few low-level functions are optimized in assembler for better speed performance. It can handle ECIES algorithm for encryption and decryption.

Library was developed using freely available software tools. We used GNU compiler from *CodeSourcery Source CodeBench Lite for ARM processors* package and *Keil μVision IDE*. The library was compiled using GNU compiler and linked to GNU and Real-View test projects. One of the successes of this work is that our library can be easily linked to projects using both kind of compilers. A big effort was made to speed up basic multi-precision multiplication algorithms for all supported processors. We are very close to the maximum performance of used MCUs architectures. We use only general purpose algorithms and we do not count special unrolled assembler codes which can be even faster.

Our future work will be concentrated in the following directions:

- Add support for general curves and support for binary field arithmetics - $GF(2^n)$.
- Add more ECC protocols.
- Even more speed could be added.
- New algorithms for multiplication, reduction, inversion can be applied.
- Add support for more processors such as ARM Cortex-M4 or Cortex-R4.
- Implement this library to soft cores on FPGA circuits.
- Use true random number generator to improve security and perform tests of the efficiency of implemented countermeasures by practical measurement using new department's laboratory.

Acknowledgment

This work is the result of the project implementation: Development of the Center of Information and Communication Technologies for Knowledge Systems (ITMS project code: 26220120030) supported by the Research & Development Operational Program funded by the ERDF.

References

[1] D. Hankerson, A. Menezes, and A. Vanstone, *Guide to Elliptic Courve Cryptography*. New York: Springer-Verlag, 2004.
[2] J. Bán, "Cryptographic library for ARM7TDMI processors," Master's thesis, Technical University of Košice, Faculty of Electrical Engineering and Informatics, Košice, 2007. [Online]. Available: http://www.kemt.fei.tuke.sk/personal/drutarovsky/students/pdfs/ban2007.pdf
[3] INVIA, "ECC Software Library IP," 2012. [Online]. Available: http://www.invia.fr/ECC-Software-Library-36.html
[4] ARM, "ARM7TDMI," 2012. [Online]. Available: http://www.arm.com/products/processors/classic/arm7/index.php
[5] ——, "Cortex-M0," 2012. [Online]. Available: http://www.arm.com/products/processors/cortex-m/cortex-m0.php
[6] ——, "Cortex-M1," 2012. [Online]. Available: http://www.arm.com/products/processors/cortex-m/cortex-m1.php
[7] ——, "Cortex-M3," 2012. [Online]. Available: http://www.arm.com/products/processors/cortex-m/cortex-m3.php
[8] Keil, "μVision," 2012. [Online]. Available: http://www.keil.com/uvision
[9] MentorGraphics, "Sourcery CodeBench Lite Edition," 2012. [Online]. Available: http://www.mentor.com/embedded-software/sourcery-tools/sourcery-codebench/editions/lite-edition/
[10] M. Brown, D. Hankerson, J. Lopez, and A. Menezes, "Software implementation of the NIST elliptic curves over prime fields," *Topics in Cryptology-CT-RSA*, 2001. [Online]. Available: http://www.dms.auburn.edu/faculty/hankerson/full-paper.pdf
[11] M. Hutter and E. Wenger, "Fast Multi-precision Multiplication for Public-Key Cryptography on Embedded Microprocessors," *Springer*, 2011. [Online]. Available: http://www.springerlink.com/content/166hl834k55r5454/fulltext.pdf
[12] M. Hutter, M. Joye, and Y. Sierra, "Memory-Constrained Implementations of Elliptic Curve Cryptography in Co-Z Coordinate Representation," *Springer*, 2011. [Online]. Available: http://www.springerlink.com/content/v48358k011x82311/fulltext.pdf

Detekcia chybnej vzorky a jej automatická korekcia pri meraní polohy sledovaného mobilného objektu

Matúš PLEVA, Anton ČIŽMÁR

Katedra elektroniky a multimediálnych telekomunikácií, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

matus.pleva@tuke.sk, anton.cizmar@tuke.sk

***Abstrakt* — Pri získavaní údajov o polohe mobilného pohybujúceho objektu sa často stáva, že chybná vzorka spôsobí, že zobrazená trajektória je fyzicky nereálna a kostrbatá. Moderné GPS systémy s presnými mapami sa snažia korigovať chybu merania polohy presunom auta na najbližšiu cestu, čo však tiež nie je ideálne ak sa auto pohybuje mimo cesty, či po ceste ktorá na mape nie je. Ďalším systémom na korekciu chyby merania, ktorá môže spôsobiť pri pomaly sa pohybujúcom či statickom objekte poskakovanie bodu na obrazovke, je takzvané statické GPS, ktoré zmení polohu zobrazeného bodu iba ak je prekročená prahová hodnota rýchlosti pohybu objektu. Navrhované algoritmy majú vylepšiť možnosti sledovania mobilných objektov operátorom na mape, pričom je možné po zadaní fyzikálnych vlastností sledovaného objektu detegovať chybné vzorky a korigovať trajektóriu s použitím Kalmanovho filtra.**

***Kľúčové slová* —Kalmanov filter, predikcia, sledovanie polohy**

I. Úvod

Pri úlohách bezpečnostných zložiek sa stáva, že je potrebné sledovať unikajúci mobilný objekt (človek, auto, motorka), ktorého poloha je snímaná mobilným snímačom polohy, ktorý získava údaje z GPS a odosiela ich cez mobilnú MANET sieť. Operátor (dispečing alebo bezpečnostná zložka s mobilným terminálom) potrebuje čo najpresnejšiu a spoľahlivú trajektóriu sledovaného objektu, pričom výpočtový systém mu môže umožniť prijať korigovanie detegovanej chybnej vzorky, čím mu uľahčí jednak rozpoznanie chybnej vzorky (to bez časových údajov o odčítanej polohe je skoro nemožné) a jednak mu ponúkne aproximáciu pravdepodobnej polohy vzhľadom na fyzikálne možnosti sledovaného objektu. Táto úloha bola motivovaná riešením pracovnej úlohy v rámci projektu 7 rámcového programu INDECT [1].

Keďže vyvinutá aplikácia má byť multiplatformová a schopná nasadenia aj na mobilné terminály, rozhodli sme sa použiť webaplikáciu, ktorej vstupom sú dodávané vzorky z GPS zariadenia sledovaného objektu s časovou značkou. Algoritmy boli najprv testované v Matlabe a následne prepísané do PHP a Javascriptu, aby mohli bežať nezávisle od použitého operačného systému a bez ďalších nárokov na programové vybavenie terminálu operátora.

II. Korekcia trasy pri zobrazení polohy

Pri zobrazení korigovaných – predikovaných polôh pomocou nižšie navrhnutých algoritmov bolo dôležité vyriešiť problém ako na predikovaných polohách vykonať dodatočnú korekciu polohy vzhľadom k novo získanej správnej polohe. A to z toho dôvodu, že pohyb objektu môže byť rôznorodý a v niektorých prípadoch dopredu nepredvídateľný. Platí, že pri viacnásobnej potrebe predikcie bodu, môže objekt sledovania prudko zmeniť smer, pričom algoritmus dostáva stále chybné dáta. Týmto „držia“ predikované body stále pôvodný smer. Je potrebné povedať, že spätná korekcia nesmie byť postavená len na jednoduchom spojení poslednej správnej polohy a najnovšej správnej polohy, s tým že sa predikované polohy len preložia na túto priamku. Je žiaduce, aby sa posledné známe vlastnosti pohybu objektu, premietli do spätnej korekcie a dotvárali túto korekciu [2].

Uveďme príklad: Auto sa pohybuje rýchlosťou 80km/h priamočiaro. Zrazu začne systém dostávať nereálne pozície tohto auta, ktoré mu poslal GPS prijímač. Systém začne generovať

predikované hodnoty, ktoré budú samozrejme vychádzať z priamočiareho pohybu auta. Ak ale auto zmení počas predikovania smer o 90° a prichádzajúce dáta budú stále nereálne, je potrebné, aby po prijatí novej správnej pozície, algoritmus zmenil trasu auta s prihliadnutím na predchádzajúce vlastnosti jeho pohybu . Dôvodom je, že pri jednoduchom spojení správnych bodov priamkou a vynesením predikovaných pozícií na ňu, by auto muselo zmeniť smer o 90° pri 80km/h okamžite, čo je v reálnej situácií samozrejme nemožné (ak nedošlo ku kolízii s hybnejším objektom – potom by sa ale už pohyb mal zastaviť či dôjsť ku poškodeniu GPS).

III. KOREKCIA TRASY KALMANOVÝM FILTROM

V roku 1960, R. E. Kalman zverejnil publikáciu, ktorá popísala rekurzívne riešenie problému lineárnej filtrácie diskrétnych dát. Od tej doby je Kalmanov filter predmetom rozsiahleho výskumu a aplikácie, najmä v oblasti samostatnej, alebo asistovanej navigácie [3] . Túto publikáciu a hlavne závery, ktoré sú v nej prezentované boli využité pri tvorbe algoritmov na predikciu polohy objektu.

Rovnice predikcie (aktualizácie v čase) pre diskrétny Kalmanov filter sú nasledovné

$$\hat{x}_k^- = A\hat{x}_{k-1} + Bu_{k-1}, \tag{1}$$

$$P_k^- = AP_{k-1}A^T + Q, \tag{2}$$

a pre korekciu (aktualizáciu novým meraním) platia tieto rovnice

$$K_k = P_k^- H^T \left(HP_k^- H^T + R\right)^{-1}, \tag{3}$$

$$\hat{x}_k = \hat{x}_k^- + K\left(z_k - H\hat{x}_k^-\right), \tag{4}$$

$$P_k = \left(I - K_k H\right)P_k^-, \tag{5}$$

z čoho vyplýva, že pri korekcii sa najprv vypočíta Kalmanov zisk K_k, potom sa generuje *a posteriori* estimácia $\hat{x}_k$. Finálnym krokom je získanie *a posteriori* kovariančnej matice P_k. V každom čase sa predchádzajúca vzorka získaná pri *a posteriori* korekcii použije ako vstup pre aktuálnu predikciu v novej *a priori* predikcii. Tento rekurzívny princíp je jednou z hlavných vlastností Kalmanovho filtra **Chyba! Nenašiel sa žiaden zdroj odkazov.**.

IV. MERANIE POLOHY

V rámci riešenia úlohy boli namerané reálne trajektóriu autobusov MHD pomocou špecializovaného GPS zariadenia GPSmap 60CSx - navigačné zariadenie, ktoré je vybavené citlivým GPS modulom. Zariadenie umožňuje okrem iného zaznamenávať prejdenú trasu a ukladať ju v rôznych časových intervaloch do formátu GPX (poloha aj časová značka pri každej odčítanej vzorke). To umožňovalo simuláciu reálnych podmienok použitia daného algoritmu.

Pri zaznamenávaní prejdených trás som si zvolil časový interval 5s. To je aj časový interval, ktorý je definovaný pre aktualizáciu pozícii v internetovom rozhraní.

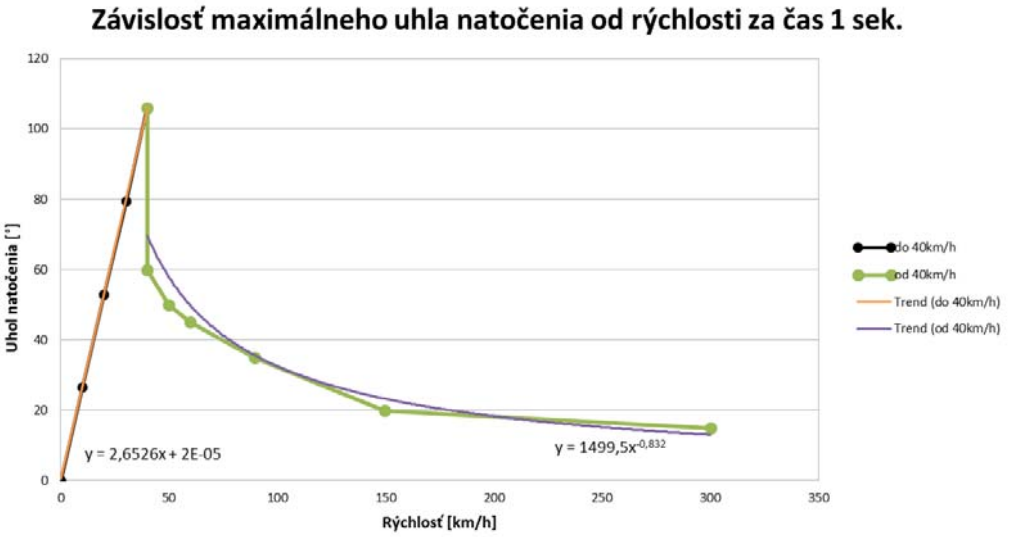


Obr. 1 GPSmap 60CSx použité pri meraniach

V. DETEKCIA CHYBNEJ VZORKY

Detekcia chybnej vzorky je založená na poznaní fyzikálnych vlastností sledovaného objektu, ktoré môžeme zhrnúť do týchto:

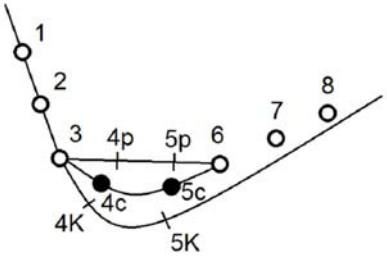
1. maximálna rýchlosť – aktuálna rýchlosť je počítaná ako vzdialenosť posledne nameraných bodov podelené časom medzi vzorkami (prípadne sa počíta priemerná rýchlosť za posledných 5 vzoriek ak sú odoberané v krátkych intervaloch, čím sa zníži nepresnosť),
2. maximálna akcelerácia – zmena rýchlosti podelená časom medzi vzorkami, pre ktoré boli dané rýchlosti vypočítané,
3. maximálna zmena smeru pri danej rýchlosti – tu je dôležité si uvedomiť, že pri malej rýchlosti je objekt obmedzený svojimi manévrovacími schopnosťami, od určitej hraničnej rýchlosti je objekt obmedzený odstredivou silou čo je vidieť na nasledujúcom obrázku.



Obr. 2 Závislosť maximálnej zmeny smeru od rýchlosti (pri aute s polomerom otočenia 6m) so zobrazením aproximovaných hraničných kriviek (trendov) [2]

VI. PRINCÍP SPÄTNEJ KOREKCIE

Dôležitou vlastnosťou systému je spätná korekcia. V princípe ide o spôsob, kedy sa predikované vzorky, ktoré vznikli pri detekcii chybných vstupných údajoch, skorigujú na základe novo prijatej správnej vzorky.



Obr. 3 Princíp spätnej korekcie

Na obrázku 3 hore môžeme vidieť princíp fungovania spätnej korekcie. Prázdne krúžky predstavujú správne odmerané a algoritmom vyhodnotené ako reálne pozície objektu. V bode tri začnú prichádzať nesprávne údaje a prediktor (algoritmus) začne predikovať hodnoty pre vzorky 4, 5 na základe posledných vzoriek (3, 2, 1, ďalšie predošlé). Táto predikcia nie je na obrázku zobrazená kvôli zachovaniu prehľadnosti obrázka. Po tom, ako algoritmus dostane ďalšiu správnu hodnotu (bod 6), vykoná spätnú korekciu pre body 4 a 5 nasledovne: Posledné správne body 3 a 6 preloží priamkou a na ňu vynesie body, ktoré boli predikované (v tomto prípade body 4p, 5p) pričom vzdialenosti medzi bodmi 3 - 4p - 5p - 6 sú rovnaké. V ďalšom kroku pomocou Kalmanovho filtra, jeho koeficientov v bode 3 (teda

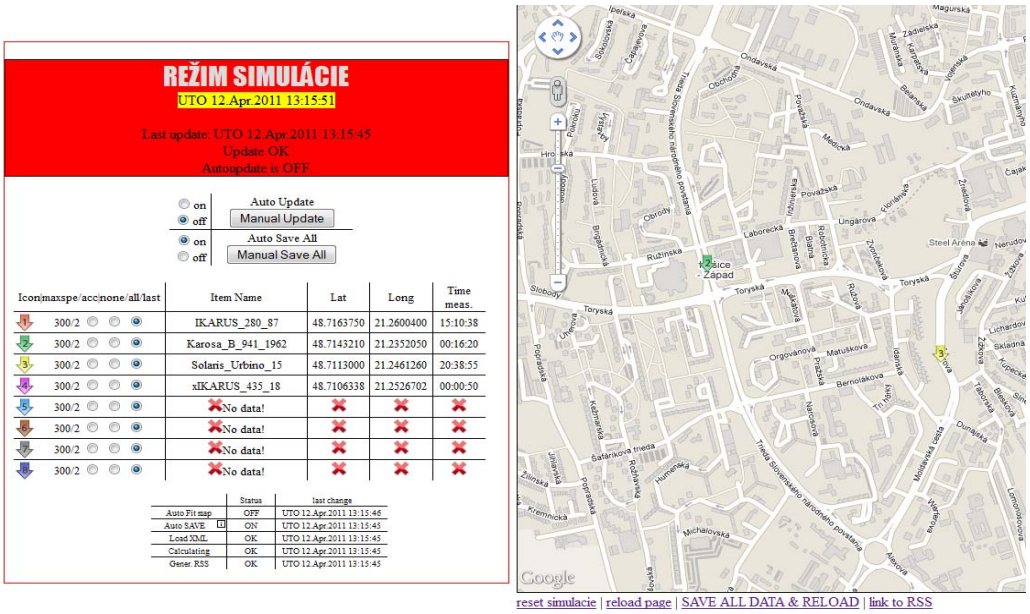
vlastností pohybu v bode 3) a na základe bodov 4p a 5p, vykoná opätovne predikciu s následnou korekciou čím vzniknú body 4K a 5K. Týmto krokom, využije vlastnosti pohybu objektu z bodu 3. V poslednom kroku spätnej korekcie, preloží body 4p s 4K a 5p s 5K priamkou, pričom v jej polovici určí body 4c a 5c, ktoré sú výsledkom spätnej korekcie [2].

VII. VYVINUTÝ MODUL PREZENTÁCIE POLOHY SLEDOVANÉHO OBJEKTU

Na obrázku nižšie je možné vidieť vyvinutý modul prezentácie polohy sledovaných objektov, kde operátor má veľké množstvo možností:

1. určiť koľko bodov zobrazených na mape (none, all, last 5)
2. určiť maximálnu rýchlosť a akceleráciu daného objektu (max spe/ max acc)
3. môže dovoliť automatickú aktualizáciu mapy pri príchode novej vzorky alebo posúvať zobrazenie trajektórie manuálne (auto/manual update)
4. môže si ukladať korigovanú trajektóriu v definovanom formáte
5. nastaviť automatické priblíženie mapy tak aby bola viditeľná celá trajektória všetkých zobrazených objektov (podľa konfigurácie v bode 1.)
6. vytvoriť RSS linku na korigované vzorky (detegované ako chybné a operátorom potvrdená ich korekcia).

Operátor tiež môže vidieť popis sledovaných objektov, ich polohu na mape (s využitím navrhnutých nových ikon), čas posledne získanej vzorky, kedy posledne došlo ku kalkulácii korekcií s využitím Kalmanovho filtra atď.



Obr. 4 Ukážka webaplikácie zobrazenia a korekcie chybných vzoriek polohy sledovaného objektu [2]

VIII. ZÁVER

V nami navrhnutom systéme bolo implementovaných viacero matematických modelov pre trajektórie mobilných objektov. Systém umožňuje po nasadení na server, ktorý prijíma dáta so senzorov polohy sledovaných objektov, veľmi jednoducho multiužívateľsky sledovať daný objekt, korigovať chybné vzorky pomocou navrhnutých predikovaných vzoriek. Výslednú trajektóriu je možné archivovať a spätne si ju prehrávať (z exportovaného XML súboru). Tento systém predstavuje dôležitý prínos pre prezentáciu polohy v reálnych aplikáciách.

POĎAKOVANIE

Táto práca bola vytvorená realizáciou projektu Rozvoj Centra informačných a komunikačných technológií pre znalostné systémy (kód ITMS projektu: 26220120030) na základe podpory operačného programu Výskum a vývoj financovaného z Európskeho fondu regionálneho rozvoja (50%) a EU ICT projektu (50%) 7. rámcového programu INDECT (FP7 - 218086).

LITERATÚRA

[1] http://www.indect-project.eu/ - stránka projektu
[2] K. Milčák, "Návrh a implementácia modulu prezentácie polohy na základe GIS dát a algoritmy pre predpovedanie postavenia sledovaných objektov," Diplomová práca, vedúci a konzultant M. Pleva, obhájená v roku 2011 na KEMT FEI TU v Košiciach, 81 strán.
[3] Weiss, H., Moore, J.B. "Improved Extended Kalman Filter Design for Passive Tracking." (1979) National Conference Publication - Institution of Engineers, Australia, (79 /4), pp. 54-58.
[4] Welch, G. – Bishop, G.: "An Introduction to the Kalman Filter" [online], Dep. of Computer Science, University of North Carolina, < http://www.cs.unc.edu/~welch/media/pdf/kalman_intro.pdf>

Využitie IP telefónie pre služby rečového komunikačného servera a jeho dodatočné zabezpečenie

Matúš PLEVA, Anton ČIŽMÁR

Katedra elektroniky a multimediálnych telekomunikácií, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

matus.pleva@tuke.sk, anton.cizmar@tuke.sk

***Abstrakt* — Z dôvodu migrácie analógových telefónnych liniek na VoIP telefóniu došlo k potrebe vyvinúť nový zvukový server (audioserver) pre služby interaktívneho komunikačného servera IRKR. Postupne sa uplatnilo viacero riešení. Jedným z nich je aj použitie ATA adaptéra, ktorý umožňuje konverziu VoIP linky na analógovú. Toto riešenie je vhodné ak je potrebné pripojiť server na propretiárny VoIP protokol, ktorý neumožňuje vývoj vlastných aplikácií. Druhým riešením je použitie otvorenej architektúry SIP protokolu a vývoj nového audioservera, ktorý bude kompatibilný s Galaxy architektúrou (cez ktorú komunikuje IRKR komunikátor) a bude sa napájať na open-source VoIP ústredňu Asterisk. Problémom pri tomto riešení bolo, že útočníci z internetu sa pokúšali získať prístup do VoIP ústredne a vykonávať hovory zdarma. Preto bol vyvinutý inteligentný adaptabilný systém na zabezpečenie ústredne, ktorý bude takisto popísaný nižšie.**

***Kľúčové slová* — VoIP, Asterisk, IRKR komunikátor, Galaxy communicator, SIP**

I. Úvod

Pri prevádzke IRKR (Inteligentné rečové komunikačné rozhranie) komunikátora [1], ktorý slúži na zisťovanie informácií o počasí a cestovných poriadkoch cez hlasové/telefónne rozhranie [2], bolo postupne navrhnuté vylepšenie služieb a ich prispôsobenie požiadavkám moderných technológií [3]. Služby sú priebežne aktualizované a vylepšované, aby sa prispôsobili zmenám v zdrojoch ich dát (webstránkam či databázam).

Služba IRKR bola v poslednej dobe takisto aktualizovaná, tak že bola doplnená o porozumenie prirodzenej reči (kl .2334), takže dialóg je jednoduchší, kratší a je možné zadať údaje celou vetou (bol nasadený aj kvalitnejší modul syntézy reči). Služba na tejto linke bola premigrovaná na propretiárnu VoIP službu s využitím ATA adaptéra. Na ďalších linkách, ktoré využívajú mobilné spojenia (telefónne čísla mobilných operátorov viď [1]) sa používajú rôzne verzie služieb a na linke SIP, ktorá je prispôsobená pre open-source riešenie beží najstaršia a najstabilnejšia služba (kl. 7817).

Zvýšenie bezpečnosti open-source VoIP služby boli vynútené častými útokmi a motivované aj prácou na riešení pracovnej úlohy v rámci projektu 7 rámcového programu INDECT [4].

Služba IRKR umožňuje aj sledovanie priebehu dialógu na rôznych použitých linkách cez monitorovaciu stránku, kde je vždy cez webaplikáciu možné sledovať aj vyťaženie systému a zoznam obsadených liniek [1].

II. Využitie propretiárnych VoIP protokolov

Služba IRKR komunikátora je multiplatformová služba založená na Galaxy komunikátore [5]. V základnej konfigurácii bol vyvinutý audioserver pre telefónnu kartu Intel Dialogic, ku ktorej sú pripojené všetky linky, ktoré služba obsluhovala. K obsluhovaným linkám pre príchodzie hovory patrili PSTN linka na Technickej Univerzite v Košiciach (kl. 2334), Skype propretiárny protokol pripojený cez SkypeMate ATA (nick: irkr_pub), mobilné linky operátorov cez GSM brány a H323 linka napojená cez požičanú ústredňu [6]. Posledne spomenutá VoIP linka bola neskôr nahradená konektivitou cez SIP sieť voľne dostupného VoIP providera napojeného cez Linksys ATA adaptér. Neskôr po otestovaní nižšie spomenutého SIP audioservera bol tento HW nahradený softvérovým audioserverom na inej báze ako pôvodný pre Intel Dialogic kartu.

Následne vznikol ale problém s migráciou telefónnej infraštruktúry PSTN na TUKE na Cisco propretiárnu VoIP sieť. Tým pádom nemohol byť použitý predtým používaný SIP adaptér, ak bola požiadavka zaistiť dostupnosť aj telefónne číslo z pôvodnej PSTN linky. Preto ako jediné riešenie bolo realizované a otestované propretiárne Cisco ATA zariadenie na obrázku nižšie (Obr. 1), ktoré momentálne obsluhuje pôvodnú PSTN klapku.

Ešte je dôležité spomenúť, že pre všetky služby napojené cez Dialogic kartu a príslušný audioserver je potrebné nakonfigurovať detektor položenia hovoru. Systém totiž nevie detegovať ukončenie hovoru používateľom priamo – iba frekvenčnou analýzou signálu na vstupe, kde hľadá tón položenia (Disconnect tone). Takže každá zmena GSM brány či ATA zariadenia vyžaduje proces nahrávania tohto tónu, jeho frekvenčnú analýzu a následné testovanie správnosti konfigurácie tohto algoritmu.



Obr. 1 Cisco ATA-187 adaptér na pripojenie 2 analógových telefónov do propretiárnej VoIP siete [7]

III. AUDIOSERVER BUDOVANÝ NA SIP PROTOKOLE

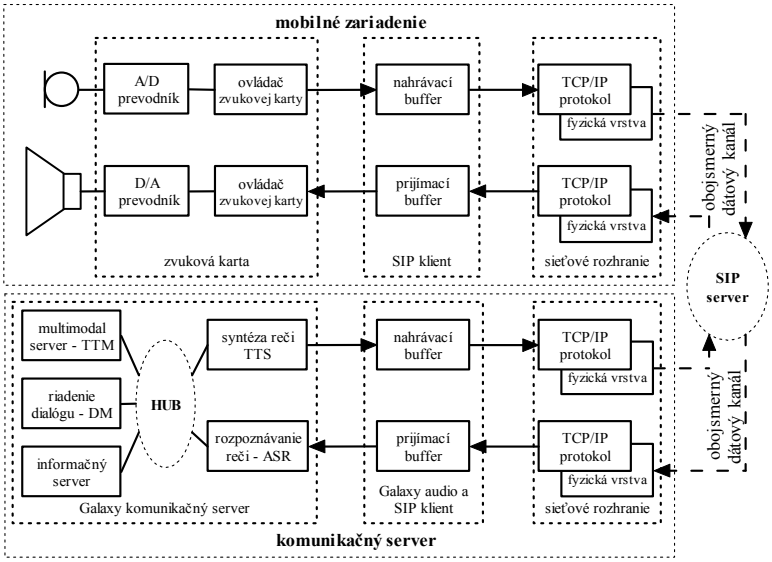
V rámci projektu Mobiltel bol vyvinutý základný jednoduchý model komunikácie cez VoIP sieť s IRKR komunikátorom, kde bola požiadavka aby z mobilných terminálov bolo možné vytvoriť jednoducho hlasový kanál [8]. Tento VoIP klient (v mobilnom termináli) pôvodne používal voľne dostupných SIP operátorov na spojenie k Linsys ATA zariadeniu pripojenému cez Dialogic kartu do IRKR komunikátora. Toto riešenie však nie je postačujúce, pretože jednak je závislé na SIP operátoroch (providers) a jednak audioserver priamo zabudovaný do Galaxy architektúry IRKR komunikátora by mohol jednoduchšie manažovať celý priebeh hovoru, jednoducho detegovať položenie hovoru, sledovať aj čísla volajúceho a robiť štatistiku pripojení, či dokonca adaptovať akustické modely pre často volajúceho zákazníka a použiť tieto adaptované modely pre jeho nasledujúce interakcie so systémom.

Podľa obrázku nižšie (Obr. 2) je zrejmé, že toto riešenie vyžaduje vývoj Galaxy kompatibilného SIP Audio servera, ktorý by bol zároveň SIP klientom pripojeným na linku SIP operátora [9]. Pri vývoji sa použil open-source SIP stack [10], ktorý bol modifikovaný a integrovaný do Galaxy servera, s tým že boli implementované funkcie prenosu signalizácie medzi SIP a Galaxy architektúrou, a hlavne hlasové/rečové kanály pre obojsmernú komunikáciu so servermi syntézy reči (TTS – Text to Speech) a automatického rozpoznávania reči (ASR – Automatic Speech Recognition).

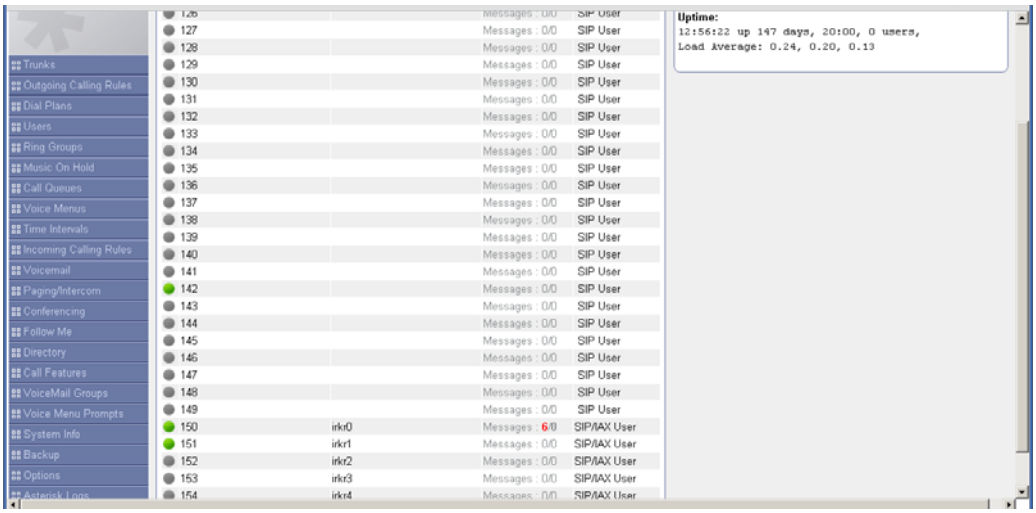
IV. OPEN-SOURCE VOIP SERVER A ROZDELENIE ZÁŤAŽE

Ďalším krokom po vyvinutí vlastného SIP klienta pre mobilné terminály a SIP audioservera (SIP klient + Galaxy server) bolo inštalácia a konfigurácia vlastného SIP servera, ktorý by nám umožňoval zasahovať do konfigurácie klientov. Konfigurácia klientov totiž umožňuje jednoducho a dynamicky vytvárať nových používateľov, prípadne dovoliť aj neautentifikovaným používateľom volať na špecifické čísla – na ktorých bežia služby IRKR komunikátora [12].

Na obrázku nižšie (Obr. 3) je možné vidieť, že po spustení viacerých SIP audioserverov je každý z nich na inej linke a tak môže byť osobitne nakonfigurovaný pre rozloženie záťaže pri volaní na IRKR služby. Napríklad je možné aby každý audioserver bol napojený na inú inštaláciu IRKR služby, alebo aspoň rôzne inštancie serverov ASR a TTS, ktoré sú najkritickejšie pri komunikácii človek-stroj.

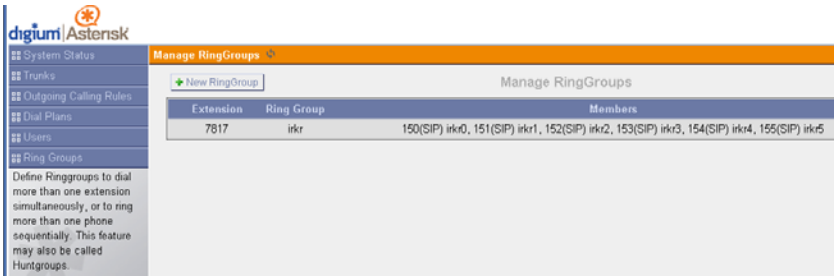


Obr. 2 Schéma prenosu zvuku pri využití SIP audioservera [11]



Obr. 3 Asterisk server – webrozhranie na sledovanie využitia liniek

Ďalšou výhodou privátneho SIP servera je, že umožňuje nastavenie volacej skupiny (Ring/Hunting Groups – viď Obr. 4 nižšie), a teda jedno telefónne číslo (kl. 7817) môže virtuálne byť smerované na viacero inštancií SIP audioservera (privátne čísla 150, 151 až 155), takže používatelia volajú len na jedno číslo a SIP server sám nájde práve voľný (zároveň funkčný – pripojený/online), prípadne dlho nepoužitý audioserver zo skupiny (používa sa aj pri službách telefónnych operátorov).



Obr. 4 Nastavenie volacej skupiny – rozdelenie záťaže na viacero fyzických liniek

V. ZVÝŠENIE BEZPEČNOSTI VOIP SERVERA AUTOMATICKOU MODIFIKÁCIOU FIREWALLU

Počas testovania a používania vlastnej inštalácie Asterisk servera [13] bol pozorovaný významný nárast chybných volaní a autentifikácií. Bolo zistené, že išlo o premyslené útoky s cieľom nájsť autentifikačné údaje do privátnej SIP siete a následne jej využitie na volanie na cudzí účet (na platené čísla).

Už počas nasadenia privátneho SIP servera bol tento nakonfigurovaný tak, aby nebolo možné volať nikam inam iba na klapky IRKR komunikátora, prípadne medzi autentifikovanými užívateľmi [12]. Neskôr sa podarilo nakonfigurovať spojenie s univerzitnou TUKE VoIP sieťou a tak boli povolené aj volania na čísla klapiek internej TUKE siete. Zároveň nášmu serveru bolo pridelených niekoľko klapiek na volanie zo siete TUKE do internej siete našej inštalácie. Vďaka tomu je možné na službu IRKR volať aj z verejnej PSTN siete a jedna klapka IRKR komunikátora v univerzitnej sieti obsluhuje niekoľko fyzických inštancií SIP audioservera a tak dokáže obslúžiť niekoľko volajúcich naraz. Zároveň však toto nastavenie neumožňuje zneužiť tento server pre volanie do platených destinácií.

Napriek tomu bolo dôležité zabezpečiť server pre neutíchajúcimi útokmi, po tom čo útočníci odhalili bežiaci SIP server a neustále skúšali rôzne kombinácie telefónnych čísiel, na ktoré by sa mohli po zadaní tajnej predvoľby dovolať (niekde je nutné pred číslo zadať 0 či inú predvoľbu).

Po úvodnom výskume bezpečnostných funkcií Asterisk servera sa nepodarilo nájsť službu, ktorá by server ochránila pred neutíchajúcimi pokusmi nad získaním kontroly či volaním zdarma. Preto bolo rozhodnuté riešiť problém systémovo na úrovni sieťového filtra, ktorý by odfiltroval všetky požiadavky na služby či autentifikáciu od IP adries, ktoré sa o to pokúsia neúspešne viac ako 5 krát. Bol použitý firewall iptables [14], ktorý sa automaticky konfiguruje na základe parsovania logovacích súborov Asterisk servera s využitím modifikovaných open-source Perl skriptov [15].

Pôvodný Perl skript bol jednak modifikovaný pre našu inštaláciu a konfiguráciu OS a firewallu. Tento skript bol periodicky spúšťaný a aktualizoval periodicky zoznam blokovaných IP adries. Po čase však bolo zistené, že neúmerne zaťažuje pamäť a procesor, pretože veľmi neefektívne pracoval s databázou IP adries nájdených v log súboroch (dokonca bol schopný úplne zahltiť pamäť servera). Po preskúmaní možností modifikácie bol vyvinutý nový nástroj založený na pôvodnej myšlienke, ktorý však používal na utriedenie IP adries a ich aktualizáciu dynamické hash polia. Vďaka tomu sa stal, program omnoho efektívnejší a pri väčších objemoch dát použiteľný aj pre priemyselné použitie.

Samozrejme rozšírený bezpečnostný modul umožňuje aj definovanie takzvaného whitelist-u, teda zoznamu IP adries, ktoré nikdy nezaradí na zoznam nežiadúcich IP adries (napríklad IP adresy internej siete TUNET, či operátorov a testovacích pracovísk).

VI. ZÁVER

V nami navrhnutom systéme bolo implementovaných viacero nových modulov a riešení, ktoré zefektívňujú beh IRKR komunikátora a rozširujú možnosti pripojenia z rôznych telekomunikačných sietí. Systém umožňuje nielen volať viacerým používateľom v tom istom čase ale dokonca aj volať na jedno číslo viacerým volajúcim naraz a inteligentný SIP server rozdelí záťaž medzi viacero serverov, či inštancií, podľa aktuálnej konfigurácie systému. Takisto bezpečnosť, ktorá je v súčasnosti najdôležitejším parametrom verejných a voľne dostupných služieb sa použitím vyvinutých riešení zvýšila na úroveň priemyselných aplikácií.

POĎAKOVANIE

Táto práca bola vytvorená realizáciou projektu Rozvoj Centra informačných a komunikačných technológií pre znalostné systémy (kód ITMS projektu: 26220120030) na základe podpory operačného programu Výskum a vývoj financovaného z Európskeho fondu regionálneho rozvoja (50%) a EU ICT projektu (50%) 7. rámcového programu INDECT (FP7 - 218086).

LITERATÚRA

[1] http://irkr.tuke.sk/ - webová stránka projektu IRKR aj so službou monitorovania priebehu dialógu
[2] Juhár, J., Čižmár, A., Rusko, M., Trnka, M., Rozinaj, G., Jarina, R. "Voice operated information system in Slovak" (2007) Computing and Informatics, 26 (6), pp. 577-603.
[3] Pleva, M., Juhár, J., Čižmár, A. "Concept of spoken dialog system based on voice over IP telephony", RTT 2008, - Bratislava, STU, 2008, 3 s., ISBN 9788022729390.
[4] http://www.indect-project.eu/ - stránka projektu INDECT
[5] http://communicator.sourceforge.net/sites/MITRE/distributions/GalaxyCommunicator/docs/manual/index.html
[6] Juhar, J., Ondas, S., Cizmar, A., Rusko, M., Rozinaj, G., Jarina, R. "Development of Slovak GALAXY/VoiceXML based spoken language dialogue system to retrieve information from the internet" (2006)

INTERSPEECH 2006 and 9th International Conference on Spoken Language Processing, INTERSPEECH 2006 - ICSLP, pp. 485-488.
[7] http://www.cisco.com/en/US/prod/collateral/voicesw/ps6788/phones/ps514/ps11026/data_sheet_c78-608596.pdf
[8] Čižmár, A., Pleva, M., Papaj, J., Doboš, L., Juhár, J. "MOBILTEL - mobile multimodal telecommunications dialogue system based on VoIP telephony" (2009) Journal of Electrical and Electronics Engineering, 2 (2), pp. 134-137.
[9] Lakatoš, L., "Softvérový SIP klient rečového dialógového systému," Diplomová práca, vedúci a konzultant M. Pleva, obhájená v roku 2010 na KEMT FEI TU v Košiciach, 55 strán.
[10] http://www.pjsip.org/ - Open Source SIP stack
[11] Juhár J. a kol. "Rečové technológie v telekomunikačných a informačných systémoch, strán 519, Equilibria, Košice, 2011, ISBN 978-80-89284-75-7.
[12] Veréb, O. "Privátna SIP sieť pre multiužívateľskú a multimodálnu aplikáciu" Diplomová práca, vedúci a konzultant M. Pleva, obhájená v roku 2010 na KEMT FEI TU v Košiciach, 81 strán.
[13] https://www.asterisk.org/
[14] https://help.ubuntu.com/community/IptablesHowTo - Basic iptables howto (official Ubuntu documentation)
[15] http://www.teamforrest.com/blog/171/asterisk-no-matching-peer-found-block/

Modely dôvery pre oportunistické siete

Ján PAPAJ, Ľubomír DOBOŠ, Peter MALIC

Katedra elektroniky a multimediálnych telekomunikácií, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

jan.papaj@tuke.sk, lubomir.dobos@tuke.sk, peter.malic@student.tuke.sk

***Abstrakt*—Tomto článok pojednáva o modeloch dôveryhodnosti využiteľné pre oportunistické mobilné siete. Tieto siete predstavujú evolúciu mobilných ad-hoc sietí, ktoré plne využívajú mobilitu zariadení. Základnou otázkou je otázka bezpečnosti. Základným prvkom bezpečnosti je poskytovanie a dosahovanie dôveryhodnosti medzi jednotlivými účastníkmi a prvkami v sieti, preto sa v tomto článku venujeme prehľadu základných modelov dôvery. Článok sumarizuje rozdelenie dôvery z pohľadu využiteľnosti v oportunistických sieťach. V dnešných moderných sieťach, kde je človek čoraz viac závislý na službách spojených s mobilitou je otázka bezpečnosti hlavným bodom, ktorým sa zaoberajú mnohé výskumné skupiny.**

***Kľúčové slová* — oportunistické siete, bezpečnosť, dôvera, sociálna dôvera, environmentálna dôvera, dôvera založená na podobnosti**

I. ÚVOD

S výhodami spojenými s mobilnými zariadeniami, mnohými scenármi sietí a aplikáciami ľudia zažínajú čoraz viac používať mobilitu a vzájomnú kooperáciu čo posúva do popredia oportunistické siete (OppNet). V otvorenom, voľne prístupnom bezdrôtovom prostredí je jednou z hlavných požiadaviek bezpečnosť v sieti. Základným prvkom bezpečnostiv OppNet s mnohými používateľmi je dôvera [1]. V decentralizovanom a rozloženom prostredí, kde nejestvuje žiadna štruktúra je dôvera ešte vyššou prioritou. V súčasnej dobe existuje niekoľko konceptov dôveryhodnosti [1], [3].Sú alternatívou kryptografie, rôznych bezpečnostných certifikátov, digitálnych podpisov s tým rozdielom, že nepotrebuje dôveryhodnú autoritu. umožňuje uzlom v sieti predpokladať budúce správanie iných uzlov. Hlavnou oblasťou použitia takéhoto zabezpečenia sú tzv. elektronické systému (e-commerce) a známe systémy podporujúce účasť na elektronických stránkach (napr. eBay). Systém zaznamenáva množstvo zobrazení stránky predajcu používateľmi. Čím vyššia je návštevnosť a dobré hodnotenie používateľov tým viac veríme predajcovi. Tento príklad predstavuje jednoduchý model dôvery, ktorý je však náchylný na rôzne druhy útokov. Dnešné systémy založené na dôvere sú omnoho komplexnejšie.

Tento článok je zameraný na modely dôvery v OppNet. Postupne je rozoberaná dôvera ako bezpečnostný koncept používaný pri návrhoch oportunistických sietí. Publikácia je rozdelená do troch základných častí. V úvode sa nachádza stručný popis dôvery, dôvody a výhody jej použitia. V kapitole II sú porovnávané dva základné pojmy dôvera a reputácia. Tieto pojmy úzko súvisia no často sa medzi sebou zamieňajú. V III kapitole sú postupne priblížené základné typy dôvery: Sociálna, environmentálna dôvera, dôvera založená na podobnosti a ako posledný bod je spomenutá kombinácia týchto typov.

II. DÔVERA VERZUS REPUTÁCIA

Dôvera spolu s reputáciou sú dva rozdielne pojmy, ktoré sa často zamieňajú. Dôvera je určitý stupeň subjektívneho očakávania. Každý používateľ podľa nej usúdi, že iná skupina (alebo zástupca skupiny) vykonajú určitú akciu. Rozhodnutie používateľa sa vykoná ešte pred tým, než môže túto akciu sledovať (alebo nezávisle na schopnosti sledovania danej akcie). Ďalej môže používateľ určiť, ako táto akcia môže ovplyvniť jeho vlastné správanie [1], [4]. Na rozdiel od dôvery, ktorá sa snaží predpovedať budúce akcie, reputácia je pasívna vlastnosť závisiaca na minulých akciách. Reputácia používateľa je dojem založený na normách správania, ktoré sú uchovávané inými používateľmi. Je založená na predošlých skúsenostiach a pozorovaniach minulých akcií. Medzi týmito dvoma vlastnosťami je samozrejme silný vzťah odkedy dôvera k používateľovi je založená na jeho dobrej reputácii [1]. Na obr. 1 je znázornený príklad reputácie medzi dvoma komunikujúcimi používateľmi.

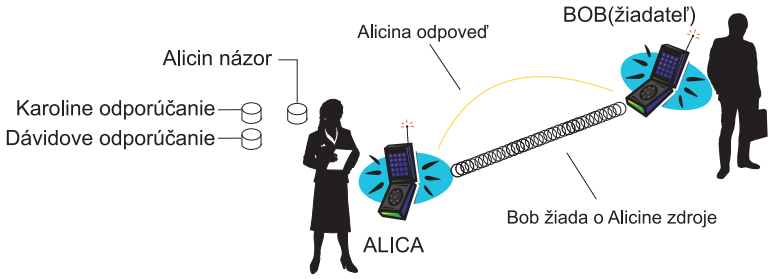


Fig. 1 Príklad použitia reputácie pri nadväzovaní novej komunikácie medzi dvomi používateľmi.

Okrem dôvery a reputácie je potrebné zohľadniť aj nasledovné funkcie [1],[2]:

1) Dôvernosť

Dôvernosť v podstate znamená vieru v to, že hlavná hodnota dosiahnutia dôvery k uzlu bola vypočítaná spoľahlivo a je správna. Odráža precíznu predpoveď budúceho správania uzla. Dôvernosť môže byť nízka ak je málo vstupných dát pre výpočet, alebo ak sú získané len na základe odporúčaní z nedôveryhodných zdrojov.

2) Riziko

Jedným z ďalších pojmov je riziko, hoci nemusí byť priamo zahrnuté do metriky daného typu dôvery. Niektoré metriky sú nastavené na zaznamenávanie benefitov. Napríklad, pri sieťach typu P2P (*peer-to-peer*) sa meria spokojnosť uzlov. Ak je vysoká, znamená to len, že riziko je relatívne nízky. Riziko je jedným z hlavných faktorov zahrnutých pri určovaní celkovej hodnoty dôvery. Je voči nej nepriamoúmerný (čím je vyššie riziko, tým nižšia je hodnota dôvery). Uzol nemusí nadviazať spojenie s iným aj keď je jeho hodnota dôvery pomerne vysoká jednoducho preto, lebo aj riziko je vyššie.

3) Dôležitosť

Dôležitosť je meradlo určujúce ako významné sú jednotlivé interakcie medzi uzlami. Interakcia medzi uzlami, ktoré majú nízku funkčnosť, ale vysokú dôležitosť je chápaná vážnejšie ako tá s vyššou funkčnosťou a menšou dôležitosťou. Dôležitá interakcia medzi uzlami môže byť asociovaná spolu s vysokou alebo s nízkou hodnotou nazývanú *trust treshold*. To sú dva extrémy riešiace ten istý problém. Ich použitie závisí na danej situácii. Časovo obmedzený uzol využije nízku hodnotu *trust treshold* a uzol s obmedzením kvality využije vysokú hodnotu. Dôležitosť je jedným z parametrov, ktoré musia byť starostlivo a uvážene vybrané pri návrhu algoritmov pre určenie celkovej hodnoty dôvery.

4) Funkčnosť

Funkčnosť je ďalší základný pojem. Problémy v dôvere môžu nastať vtedy, ak je táto vlastnosť spojená s interakciami. Napríklad, v elektronickom systéme, ako napríklad internetový obchod sa hodnota dôvery v predajcu výrazne zníži na základe množstva podvodných interakcií. Používatelia potom prestávajú nakupovať od daného predajcu. Škodlivé uzly sú v tomto prípade priamo v systéme, keďže sú zaregistrovaní a majú vytvorený užívateľský účet. Bežná autentifikácia je tu preto neúčinná a zbytočná. Výpočet dôvery zahrňujúci hodnotu funkčnosti pomáha brániť útokom takéhoto typu. Táto hodnota taktiež diferencuje prístup k zdrojom tým, že limituje počet používateľov a zvyšuje tak kontrolu nad celým systémom.

III. FORMY DÔVERY V OPORTUNISTICKÝCH SIEŤACH

Dôveru môžeme chápať vo viacerých formách. Napríklad, dôverujeme nášmu stredoškolskému kamarátovi, pretože sme s ním strávili veľa času, poznáme jeho pôsobenie a okolie tak isto ako jeho charakter a zvyky a podľa toho sme schopní posúdiť čoho je schopný. Takisto veríme študentovi, ktorého nepoznáme, ale stretávali sme ho na mnohých hodinách. V hlavnom zmysle je to pre to, že poznáme jeho zameranie a činnosť. Dôležitejšie však je to, že ho očakávame aj na nasledujúcej hodine. Okrem toho, dôverujeme aj kritikovi filmu, ktorého sme pred tým nikdy v živote nestretli, ale ktorý poskytol názor podobný nášmu a očakávame od neho podobné rady aj v budúcnosti. Nižšie sú popísané rôzne druhy dôvery aplikované na rôzne požiadavky. Najprv to bude sociálna dôvera používaná pri vedomom definovaní väzieb priateľstva a potvrdzujúca fakt, že mobilita pomáha pri bezpečnosti. Na druhej strane je tu dôvera prostredia odvodená od znalosti okolitých uzlov, využívajúca komplexné sieťové štruktúry, vyplývajúce zo sociálneho charakteru oportunistických sietí.

Ďalším typom dôvery je dôvera založená na podobnosti. Zvažuje korelácie hodnotení, alebo názorov, ktoré sú medzi jednotlivými používateľmi podobné. Poslednou spomenutým typom je

kombinácia jednotlivých foriem dôvery. Všetky tieto typy dôvery sú medzi sebou ortogonálne a môžu byť hodnotené nezávisle. Formy dôvery môžu byť vyvažované a navzájom kombinované na základe rôznych požiadaviek[2].

A. Sociálna dôvera

Základom sociálnej dôvery je napred vedome zvolená väzba priateľstva. S výhodou mobility zariadení môže byť bezpečnosť a spoľahlivosť väzieb priateľov ľahko dosiahnuteľná rôznymi formami bezpečného párovania. Táto forma má samozrejme okrem výhod aj svoje nevýhody. Počas nadviazania priateľstva je za povolením užívateľa nejaká ľudská entita, ktorá rozhodne o dôvere k danému užívateľovi a nadviazaní spojenia [2]. Na druhej strane dôvera založená na týchto sociálnych vzťahoch predpokladá presnú znalosť o používateľovi a o jeho akciách. Uzly v OppNet, ktoré sú vo vzájomnom dosahu a nadviazali spojenie si vymieňajú navzájom svoje zoznamy priateľstva a dopĺňajú *graf priateľstva*.

Graf priateľstva je špeciálny graf organizovaný do úrovní (L_d) tvorených súborom uzlov s rovnakou vzdialenosťou od lokálneho uzla. Hraničné uzly existujú len medzi uzlami v usporiadaných úrovniach. Tento graf sa realizuje pomocou mierne upraveného BFS algoritmu (*Breadth-First Search algorithm*)[2]. Táto modifikácia je nutná preto, aby bolo možné povoliť hraničné uzly medzi jednotlivými úrovniam. Algoritmus vypočíta hodnotu dôvery v rozsahu [0, 1]. Všetkým priamym priateľom pridelí hodnotu dôvery 1. Hodnota sa postupne znižuje s každým ďalším preskokom a závisí na kvalite pripojenia daného uzla. Jej zníženie je rýchlejšie ak existuje veľa väzieb priateľstva. Hodnota dôvery sa neznižuje vôbec ak existuje len jeden uzol v každej úrovni Ld. Na obr. 2 je príklad grafu priateľstva s jednotlivými hodnotami dôvery a spojeniami medzi uzlami v rôznych úrovniach.

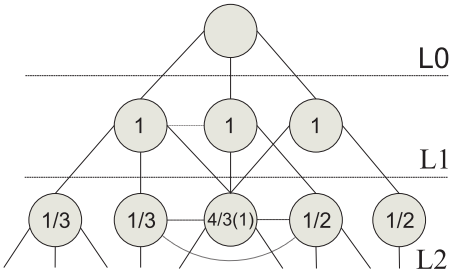


Fig. 2 Graf priateľstva organizovaný do levelov Ld s hodnotami dôvery v každom uzle[2].

B. Environmentálna dôvera

V každodennom živote existujú určité osoby, s ktorými si zdieľame svoj životný priestor, alebo rôzne aktivity. Napríklad, v bytových domoch stretávame pravidelne každé ráno tých istých susedov, v kancelárii spolupracujeme väčšinou s tými istými kolegami a každý utorok a štvrtok sa istými ľuďmi stretávame v posilňovni. Algoritmus detegovania komunity sa pokúša identifikovať komunity starostlivou analýzou a pozorovaním prostredia a entít v oblasti v danom čase. Tieto komunity nemusia byť sociálne, aj keď sú tak nazývané, pretože členovia daného spoločenstva o sebe nemusia vedieť a nemusia mať vytvorené väzby priateľstva. Entity v komunite len zdieľajú určitý priestor v danom čase. Priradenie do spoločenstva nezaručuje, že určitý subjekt má v skutočnosti identitu, pod ktorou vystupuje, čo znamená, že táto forma dôvery nie je tak bezpečná ako sociálna dôvera, kde je potrebná znalosť identity subjektu na vytvorenie väzby priateľstva. Napriek tomu, presné množstvo dôvery k známemu cudzincovi (v skutočnosti ho nepoznáme nič o ňom nevieme no pravidelne ho stretávame) môžeme určiť. Identita nie je krátkodobá, čo môže byť užitočné proti útoku typu Sybil a výrazne to zvyšuje úsilie potrebné na vykonanie útoku. Žiadna interakcia medzi entitami prostredia je výhodou oproti metóde vytvárania väzieb priateľstva medzí používateľmi. Dôležitým aspektom mechanizmu detekcie komunít je jeho ľahké ovplyvnenie a manipulácia, preto je potrebná zvýšená pozornosť pri zabezpečovaní jeho odolnosti.

Jedným zo spôsobov analýzy prostredia je práve pomocou detegovania rozličných foriem komunity. Komunita môže byť popísaná ako dobre spojené spoločenstvo entít. Analýzou trvania kontaktu a/alebo frekvencie kontaktu medzi jednotlivými uzlami a pomocou zdieľania týchto informácií medzi nimi môžeme určiť komunitu do ktorej každý uzol patrí. V tejto forme dôvery sú navrhnuté tri algoritmy: *Simple*, *k-Clique* a *Modularity* [3], [4], [5]. Každý z týchto algoritmov potrebuje udržiavať a vymieňať minimálne súbor známych členov a členov komunity. Používateľ je pridávaný do súboru známych členov po nastavení presnej hodnoty jeho familiárnosti (akumulovaný čas výskytu v blízkosti alebo frekvencia výskytu v blízkosti používateľa). Pre vybudovanie komunity používa každý algoritmus svoju vlastnú stratégiu. Algoritmus *Simple* porovnáva prostredia dvoch uzlov s cieľom rozhodnúť, či sa navzájom

pridajú do ich vlastných komunít. Algoritmus *K-Clique* hľadá slučky určitých veľkostí, do ktorých uzol patrí a algoritmus *Modularity* sa snaží nájsť komunitu optimalizovaním modularity ako vo väčšine algoritmov s detekciou komunity [3],[4],[5].

C. Dôvera založená na podobnosti

Dôvera môže byť založená na podobnosti na základe osobitého názoru jednotlivých používateľov. Táto metóda je zvyčajne kombinovaná s ďalšími systémami najčastejšie s využitím reputácie. Dôvera v takomto systéme je hodnotená na základe porovnávania odporúčaní od jednotlivých uzlov s priamymi skúsenosťami. Dodatočne bola navrhnutá ďalšia schéma tejto dôvery pre siete typu P2P. Tento spôsob hodnotenia dôvery je veľmi komplikovaný, obzvlášť pre prípad oportunistických sietí [2].

D. Kombinácia dôvery

Veľmi dôležitým aspektom výberu dôvery je cieľ a oblasť jej použitia. Každá forma dôvery má iné vlastnosti. Z toho vyplýva, že nie každý typ je vhodný na použitie vo vybranej oblasti a je nutné zvážiť rôzne požiadavky. Zatiaľ čo sociálna dôvera je najčastejšie zameraná na overenie používateľov, dôvera na základe podobnosti uzlov je užitočná pri získavaní hodnotných názorov o ostatných uzloch[6]. Adaptácia na rôzne požiadavky a potreby môže byť popísaná spojením rôznych typov dôvery. Interakcia uzlov na základe kombinovanej dôvery môže byť popísaná nasledovne [7]:

$$\mathrm{w}_s * t_s + w_e * t_e \geq th_{lu} \tag{1}$$

kde w_s a w_e sú váhy jednotlivých typov (s-sociálna, e-environmentálna) a t_s, t_e sú jednotlivé hodnoty dôvery. th_{lu} predstavuje prah použitia.

IV. ZÁVER

Oportunistické siete novej generácie prinášajú množstvo výhod spojených s mobilitou zariadení prenášaných ľuďmi. Tieto siete menia doterajší pohľad na komunikáciu medzi mobilnými zariadeniami prostredníctvom priamej skokovej komunikácie medzi nimi aj v rozvojových prostrediach kde nie je vystavaná centralizovaná infraštruktúra. Používatelia sú zapojení do kolektívnej komunikácie s prostredím. Je možné využívať rôzne aplikácie poskytujúce zážitok zo skutočných sociálnych sietí prostredníctvom digitálnych technológií. V týchto aplikáciách však dochádza k publikovaniu rôzneho obsahu používateľmi. V takomto otvorenom prostredí, bez centralizovaného riadenia sú dôležitými aspektmi otázky bezpečnosti a smerovania dát medzi komunikujúcimi používateľmi. Sú navrhované mnohé metódy zabezpečenia a rôzne smerovacie algoritmy, ktoré neobmedzujú mobilitu účastníkov. V týchto druhoch sietí nie je možné centralizované zabezpečenie pretože zariadenia komunikujú navzájom medzi sebou. Sú využívané rôzne modely dôvery medzi zariadeniami. Tieto modely sú založené na rôznych kritériách a posudzujú rôzne vlastnosti sietí a užívateľov. Sú nezávislé a môžu byť navzájom kombinované, čím je možné dosiahnutie požadovaných kritérií, autentifikáciu a zabezpečenie komunikácie v rôznych prostrediach.

POĎAKOVANIE

Táto práca bola vytvorená realizáciou projektu Výskum modulov pre inteligentné robotické systémy (kód ITMS projektu: 26220220141) na základe podpory operačného programu Výskum a vývoj, financovaného z Európskeho fondu regionálneho rozvoja (50 %) a projektom INDECT (FP7-No.218086) (50%).

LITERATÚRA

[1] N. K. Kulesza, “User behaviour and security in opportunistic networks”. Londýn : ImperialCollegeofLondon Department ofcomputing, 2009, pp. 89. Dostupné na internete: http://www.doc.ic.ac.uk/teaching/distinguished-projects/2009/n.kulesza.pdf

[2] C. Hausl , J. Hagenauer, „Relay communication with hierarchical modulation“, IEEE CommunicationsLetters, vol. 11, pp. 64-66, Jan., 2007

[3] S. Trifunovic, F. Legendre,” Trust in opportunisticnetworks”. ComputerEngineering and NetworksLaboratory, 2010. Švajčiarsko, Dostupné na internete: ftp://ftp.tik.ee.ethz.ch/pub/publications/TIK-Report-318.pdf

[4] ChingLin and VijayVaradharajan, “Trust based risk managementfordistributedsystemsecurity - a new approach”. V ARES 2006: ProceedingsoftheFirstInternationalConference on Availability, Reliability and Security, pp. 6 -13, Washington, DC, USA, 2006. IEEE Computer Society.

[5] M. E. J. Newman, “Fast algorithm for detecting community structure in networks”. PhysicalReview E, Vol. 69, pp. 066133, 2004.

[6] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, “Fast unfoldingof communities in large networks”. Journal of Statistical Mechanics: Theory and Experiment, Vol. 2008, No. 10, 2008.

[7] A. Clauset, “Findin glocal community structure in networks”, Physical Review E,Vol. 72, No. 2, 2005.

Teória hier a jej využitie pre bezpečnosť v kooperatívnych sieťach

Ján PAPAJ, Ľubomír DOBOŠ, Tomáš LOZIAK

Katedra elektroniky a multimediálnych telekomunikácií, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

jan.papaj@tuke.sk, lubomir.dobos@tuke.sk, tomas.loziak@student.tuke.sk

***Abstrakt*—V súčasnej dobe je veľmi aktuálna problematika bezpečnosti v moderných mobilných sieťach ako aj v kooperatívnych sieťach. Tento typ siete pri svojom multiskokovom prenose využíva tranzitné uzly kvôli zvýšeniu kvality spojenia, zvýšeniu prenosového výkonu a zníženiu spotreby energie. Avšak realizácia spojenia v kooperatívnej medzi rôznymi zariadeniami kdekoľvek na svete nie je možná bez efektívneho mechanizmu na stimuláciu kooperácie jednotlivých uzlov. Tieto mechanizmy sú realizovateľné vďaka modelu teórie hier. V neprítomnosti centrálneho riadenia sú taktiež potrebné nové spôsoby mechanizmov bezpečnosti a ochrany súkromia vyžadujúce si medzivrstvový prístup. Tento článok zároveň poskytuje stručný pohľad do problematiky bezpečnosti pre kooperatívne siete.**

***Kľúčové slová* — Kooperatívne siete, bezpečnosť, teória hier, modely teórie hier**

I. ÚVOD

Kooperatívny prenos dát umožňuje výrazne zvýšenie kvality prenosu medzi zdrojovým a cieľovým uzlom pri využití tranzitných uzlov. Takéto kooperatívne spojenia môžu byť použité ako základný stavebný blok pri prenose medzi koncovými užívateľmi vo viacskokovej sieti. Podľa spôsobu prenosu cez tranzitný uzol rozoznávame fixné kooperujúce protokoly *Dekóduj a Prepošli* (*Decode and Forward - DF*) a *Zosilni a Prepošli* (*Amplify and Forward - AF*) [1], alebo adaptívne medzi ktoré patrí najpoužívanejší *Selektívny DF* [2].

Metóda integrujúca kooperáciu do kódového kanála sa nazýva *Kódová kooperácia.* Aby mohli byť v kooperujúcej sieti využité tranzitné uzly a kooperujúce protokoly na fyzickej vrstve, museli byť vyvinuté smerovacie algoritmy, ktoré ich v smerovaní dokážu zohľadniť a využiť. Takéto algoritmy sa nazývajú *heuristické* a patria medzi ne napríklad algoritmus *Kooperácie pozdĺž energeticky najvýhodnejšej nekooperatívnej cesty* (Cooperation Along the Minimum Energy Non-Cooperative Path - CAN) [3], alebo algoritmus *Najkratšej kooperatívnej cesty* (Cooperative Shortest Path - CSP)[4].

Avšak k zavedeniu kooperujúcich sietí je ešte potrebné preskúmať bezpečnosť v týchto sieťach a zaviesť potrebné bezpečnostné mechanizmy, aby mohli byť tieto siete využívané [14].

Tento článok sa zaoberá bezpečnosťou pre mobilné kooperujúce siete. Problematika je rozdelená do štyroch základných častí. V úvode sú priblížené kooperujúce siete a ich teoretické základy. V kapitole II sú špecifikované základné bezpečnostné problémy, ktoré sa začínajú javiť v týchto sieťach. V kapitole III je predstavený základný model teórie hier, ktorý je využívaný v mechanizmoch na podporu kooperácie, a tieto sú následne stručne popísané v kapitole IV. Mechanizmy na podporu kooperácie slúžia na posilnenie kooperácie a vylúčení sebeckých uzlov zo siete. V kapitole V je popísaný stručné využitie medzivrstvového návrhu bezpečnostného mechanizmu, ktorý vznikol skĺbením mechanizmov na bezpečnosť v typických MANET s mechanizmami špecifickými pre kooperujúcu sieť.

S týmto cieľom je najprv popísaný bezpečný prenos cez jeden tranzitný uzol pomocou vyhodnotenia vzájomnej korelácie symbolov, následne aj bezpečný prenos cez viacero tranzitných uzlov vďaka stopovacím symbolom. Tento systém by mal slúžiť ako základ pre budúci návrh medzivrstvového modelu chrániaceho kooperujúcu sieť [5].

II. KONCEPT TEÓRIE HIER V KOOPERATÍVNYCH SIEŤACH

Ak sa tranzitné uzly nepoddajú kooperácii, alebo škodlivo prenášajú skreslené správy za účelom rušenia komunikácie v priamom kanáli, výkon detekcie na cieľovom uzle môže významne klesnúť. V takomto prípade je pre používateľov výhodnejšie komunikovať bez kooperácie.

V niektorých prípadoch tranzitné uzly nemusia byť škodlivé, ale iba sebecky uchovávajú svoje prenosové prostriedky pre vlastné použitie. Toto môže spôsobiť zrútenie kooperácie dovtedy, kým ostatné uzly nemajú podnet ku kooperácii. Pre cieľový uzol je preto dôležité odlíšiť škodlivé uzly od sebeckých, aby mohli vykonať patričné opatrenia na obnovení plnej a bezpečnej kooperácie. V mobilných sieťach sa rozlišuje správanie uzlov [5], [10]:

A. *Sebecké správanie* (Selfishrelaying) - Prípad v ktorom uzly nevykladajú požadované množstvo prostriedkov na posielanie dát od zdrojového uzla. Namiesto toho šetria energiu pre vlastné prenosy.

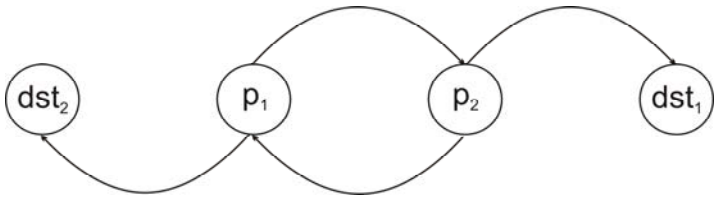
B. *Nesprávne informácie o stave siete* (False information feedback) - Prípad v ktorom sú ostaným uzlom v sieti odosielané nesprávne informácie typu smerovacej cesty alebo nákladov na prenos. Tieto útoky spôsobujú škodlivé účinky pri koordinácii viacerých užívateľov a smerovanie medzi nimi.

C. *Škodlivé preposielanie* (Malicious forwarding) - Prípad, kedy sú signál, alebo správa skreslené škodlivým uzlom a následne preposielané namiesto pôvodného obsahu. Takéto útoky spôsobujú interferenciu v cieľovom uzle a zhoršenú detekciu signálu.

D. *Odpočúvanie* (Eavesdropping) -Prípad v ktorom útočníci zámerne odpočúvajú správy, ktoré nie sú určené pre nich. Porušujú tým súkromie ostatných uzlov.

A. MODEL TEÓRIE HIER

Medzi najznámejšie hry, ktoré sa využívajú v kooperujúcich sieťach, patrí hra nazývaná FD (*Forwarder´sDilemma – Dilema Preposielania*), kde vystupujú dva zariadenia, ktorý sú označené ako hráči p_1 a p_2. Pri hre obidvaja hráči chcú odoslať komunikáciu k cieľu jednotlivo v každom časovom slote využijúc druhého hráča ako *preposielateľa*. Pri hre predpokladáme, že komunikácia medzi hráčom a jeho prijímateľom je možná len ak ju bude druhý hráč preposielať. Dilema Preposiealania je zobrazená na obr.1[6], [8].



Obr. 1 Princíp hry Dilema Preposielania a jej základný algoritmus.

Ak hráč p_1 preposiela správu pre hráča p_2, hráč p_1 má *náklady* (*cost*) na prenos v rozsahu *0 < C << 1*, ktoré reprezentujú energiu a výpočty (*computations*) potrebné na preposielanie. Umožní tým prenos medzi p_2 a dst_2, čím dáva hráčovi p_2 benefit *1*. Výnos je potom rozdiel medzi benefitom a nákladmi. Pri predpoklade, že je táto hra symetrická, rovnaké uvažovanie je aplikované aj pri preposielacom ťahu na hráča p_2.

Tabuľka 1
Dilema Preposielania

p_1	p_2		
		P	Z
	P	(1 – C, 1 – C)	(- C, 1)
	Z	(1, - C)	(0, 0)

Hlavná dilema je teda nasledujúca: Každého hráča láka nepreposlať komunikáciu, ktorú by mal preposielať, pretože to môže ušetriť jeho vlastné zdroje. Avšak ak aj druhý hráč má

rovnaké dôvody, potom komunikácia ktorú chce odoslať prvý hráč bude takisto zahodená. A preto by bolo lepšie ak by si vzájomne preposielali komunikáciu toho druhého.

Táto dilema bude rozobraná do strategickej podoby. Odohráva sa len v jednom časovom slote a hráči môžu komunikáciu preposlať (P) alebo zahodiť (Z). Toto rozhodnutie reprezentuje stratégiu hráča. Vhodnou reprezentáciou hier v strategickej forme sú matice. Takto môžeme znázorniť Dilemu preposielania v tabuľke č.1 [6].

Hra $G = (N, S, U)$ môže byť definovaná prostredníctvom *strategickej formy* s nasledujúcimi troma premennými. N je množina hráčov. V našich príkladoch uvažujeme dvoch hráčov $p_1, p_2 \in N$, ale definície budú uvádzané pre akýkoľvek počet hráčov. Pre lepšiu znázorniteľnosť bude vymedzená táto množina indexom $-i$, čiže všetci hráči patria N okrem i. Títo hráči sú označovaní ako *oponenti* hráčovi i. V našej hre má hráč i jedného oponenta j. S_i označuje *úplný strategický priestor* hráča i. To znamená, že táto stratégia dáva nulovú pravdepodobnosť všetkým ťahom, okrem jediného. Je potrebné definovať množinu strategických priestorov $S = S_1 \times ... \times S_{|N|}$. Úplný strategický priestor oponentov hráča i je daný ako $S_{-i} = S / S_i$. Súbor zvolených stratégií predstavuje *strategický profil* $s = \{s_1, s_2\}$. V našich dilemách majú obaja hráči rovnaký strategický priestor $S_1 = S_2$. Strategický profil oponentov teda bude $s_{-i} = s_j \in S$. Výnos $u_i(s)$ vyčísľuje výsledok hry hráča *i* so strategickým profilom s. V našom prípade $U = \{u_1(s), u_2(s)\}$. Taktiež je dôležité ustanoviť, že hra prebieha s *úplnou informovanosťou.*

Následne je hra definovaná v strategickom tvare a je potrebné ju vyriešiť. Vyriešením hry sa myslí predvídanie stratégie každého hráča, posudzujúc na základe informácií o hre a s predpokladom racionálneho správania sa hráča. Na vyriešenie hry existuje niekoľko spôsobov, najjednoduchší z nich na to využíva *prísnu dominanciu*[6]. Stratégia $s_i^{'}$ hráča i je pod prísnou dominanciou stratégie s_i ak platí [7],[8]:

$$u_i(s_i^{'}, s_{-i}) < u_i(s_i, s_{-i}), \forall s_{-i} \in S_{-i}$$

Riešenie hry (tab. 1) pomocou prísnej dominancie bude nasledujúce. Z pohľadu hráča p_1, preňho je P stratégia pod prísnou dominanciou Z stratégie. To znamená, že z matice možno vyradiť prvý riadok, pretože racionálny hráč takú stratégiu nikdy nezvolí. Z rovnakého dôvodu, avšak z pohľadu hráča p_2, nasleduje vyradenie prvého stĺpca matice. Výsledkom je riešenie hry (Z,Z) a následné výnosy činia (0,0). Tento výsledok sa zdá byť trochu paradoxný, keďže riešenie (P,P) by dospelo k lepším výnosom pre každého hráča, ale bez podporného mechanizmu spolu uzly kooperovať nebudú[7].

B. TEÓRIA HIER V SYSTÉMOCH STIMULUJÚCICH KOOPERÁCIU

Ako je ukazuje model teórie hier v FD, samotné uzly nebudú kooperovať s cieľom ušetriť energiu pre vlastné prenosy. Aby sa situácia v matici zmenila a jednotlivé uzly boli prinútené kooperovať, je potrebné zaviesť mechanizmy podporujúce kooperáciu. Tieto systémy využívajú model teórie hier pre analýzu správania sa tranzitných uzlov[5],[11].

MODEL TEÓRIE HIER PRE SYSTÉM ZALOŽENÝ NA REPUTÁCII

V takomto systéme sa zavádza pojem medzná hodnota *reputácie,* nevyhnutný na oddelenie sebeckých uzlov od kooperujúcich. Ak sa susedné uzly podieľajú na kooperatívnom prenose, reputácia týchto uzlov rastie v uzloch, ktoré to sledujú. Naopak ich reputácia klesá, ak sa na kooperujúcich prenosoch nepodieľajú. Ak reputácia sebeckých uzlov klesne pod medznú hodnotu, smerovacie požiadavky týchto uzlov budú zamietnuté všetkými ostatnými uzlami. Problémom začína byť situácia keď každý uzol pozná svoju reputáciu a v situáciách keď ju má vysokú začne pakety zahadzovať, kým sa nepriblíži k medznej hodnote reputácie, čím si ušetrí energiu.

MODEL TEÓRIE HIER PRE SYSTÉM ZALOŽENÝ NA VYPLÁCANÍ

V tomto systéme sú ako stimul ku kooperácii použité *virtuálne peniaze*. Uzol, ktorý preposiela komunikáciu je odmenený virtuálnymi peniazmi (náklady), ktoré budú odpočítané zdrojovému uzlu. Požiadavky na prenos komunikácie budú odmietnuté, ak ich zdrojový uzol

nemá čím zaplatiť.Aj v tomto modeli nastáva nežiaduci stav keď uzol nahromadí veľkú hodnotu virtuálnych peňazí, môže začať zahadzovať komunikáciu až kým mu nedôjdu peniaze.

MODEL TEÓRIE HIER PRE ARM SYSTÉM

Stimulujúca stratégia v ARM systéme v sebe kombinuje oba spomínané systémy. Systém založený na reputácii je použitý na posúdenie stupňa kooperácie každého uzla, podľa ktorého sú vyplácané odmeny v podobe virtuálnych peňazí za preposielanie komunikácie. Uzol s vyššou reputáciou platí menšie ceny. Ak hodnota reputácie uzla klesne pod medznú hodnotu, tento uzol bude preložený na "čiernu listinu" nezáležiac od toho aký je bohatý (koľko virtuálnych peňazí vlastní. Uzlom z čiernej listiny budú následne odmietnuté akékoľvek požiadavky na prenos informácií.

III. BEZPEČNOSŤ V KOOPERUJÚCEJ SIETI

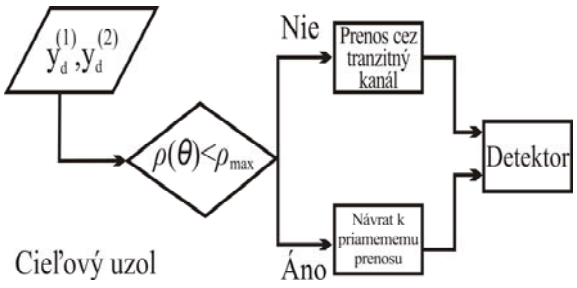
Pre poskytnutie zabezpečeného existuje niekoľko stratégií na zaistenie spoľahlivého prenosu v kooperujúcich sieťach obsahujúcich používateľov so škodlivým správaním [10],[11],[12],[13]. Pre siete obsahujúce škodlivé uzly je možné použitie dohľadových metód (watchdog a pathrater) na identifikovanie spoľahlivej smerovacej cesty od zdrojového k cieľovému uzlu, bez narušenia sebeckými či škodlivými uzlami.

Metóda Watchdog rozoznáva oba tieto typy tranzitných uzlov vďaka neustálemu monitorovaniu prenosových aktivít svojich susedov [15]. Sebecký uzol, ktorý neodošle prijatú správu alebo škodlivý uzol, ktorý neprenesie správny paket je detegovaný svojimi susedmi, najmä uzlom predošlého skoku, a označený ako uzol s nežiaducim správaním. Následne po príchode smerovacej požiadavky, patharater vypočítava cenu každej cesty posudzujúc cenu prenosu spoločne so spoľahlivosťou každého uzla na smerovacej ceste.

Väčšina bezpečnostných techník pre obvyklé ad hoc siete vyžaduje monitorovanie každého samostatného prenosu a následne bojujú proti škodlivým používateľom. Avšak v kooperujúcich systémoch je správa zvyčajne preposielaná súbežne viacerými tranzitnými uzlami a teda nie je jednoduché prekontrolovať individuálne správanie jedného tranzitného uzla. Dokonca pri detekcii zlyhaní prenosu kvôli škodlivému správaniu, je zložité identifikovať škodlivý uzol uprostred skupiny kooperatívnych používateľov použitím bežných bezpečnostný praktík spomenutých vyššie. Použitím týchto techník vyššej vrstvy, všetky tranzitné uzly zahrnuté v nevydarenom prenose môžu byť považované za uzly s nežiaducim správaním, aj keď prenos nezlyhal ich pričinením. Kým dôveryhodné uzly budú vylúčené z kooperatívneho prenosu zisk z diverzity bude znížený. Preto je potrebná spolupráca viacerých vrstiev - medzivrstvový návrh. V ďalších podkapitolách budú preto predstavené bezpečnostné techniky na najnižších vrstvách, ktoré v spojení s technikami na sieťovej úrovni tvoria spomínaný medzivrstvový návrh[12].

A. Bezpečnosť v prenose cez jeden tranzitný uzol

Pri predpoklade vynikajúceho stavu kanálu v cieli prenosu, signály prijaté zo zdrojového uzla a tranzitného uzla môžu byť koherentne kombinované k vykonaniu detekcie. V tomto prípade prenosu cez jeden tranzitný uzol, môže byť táto detekčná metóda využívaná popri ostatných bezpečnostných metódach k zvýšeniu bezpečnosti prenosu cez danú sieť. Pri použití kooperačného protokolu DF, čiže dvojfázovej kooperačnej schémy, cieľový uzol prijme dve kópie zdrojového signálu, jeden prenášaný priamo od zdrojového uzla, druhý preposlaný cez tranzitný uzol. Z toho vyplýva, že ak sa tranzitný uzol podriadi kooperatívnej stratégii, vysielané signály budú vysoko korelačné. A teda po preskúmaní korelácie medzi obidvoma signálmi, cieľový uzol dokáže rozhodnúť či sa tranzitný uzol správa nežiaduco, alebo nie.



Obr. 2 Detekcia uzla s nežiaducim správaním v cieľom uzle.

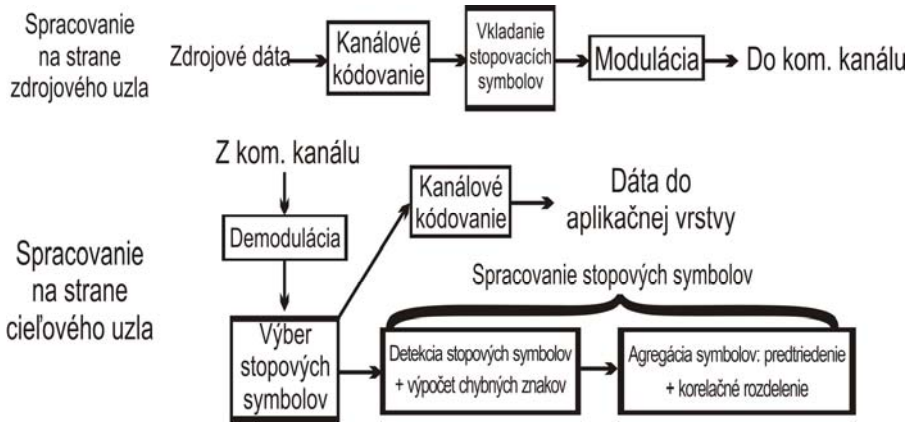
Obr.2 prezentuje diagram detekcie v cieľovom uzle, kde $y_d^{(1)}$ predstavuje prijatý signál v prvej fáze, $y_d^{(2)}$ signál prijatý v druhej fáze, ρje miera korelácie prijatých symbolov a ρ_{max} maximálna miera korelácie pri nežiaducom správaní škodlivého uzla. Cieľový uzol po prijatí oboch signálov preverí stav kooperácie tranzitného uzla porovnaním miery korelácie ρ s hodnotouρ_{max} apodľa toho rozhodne v akom móde bude prebiehať nasledujúca komunikácia. Je dôležité podotknúť, že k vyhodnoteniu správnej korelácie prijatých symbolov, cieľový uzol potrebuje prijať viacero symbolov za určitý časový úsek až následne vypočíta priemernú koreláciu prijatých symbolov. Tento proces je možné zovšeobecniť na prípad prenosu cez viaceré tranzitné uzly, obzvlášť ak predpokladáme prenosy cez ortogonálne kanály [12].

B. Bezpečnosť v prenose cez viacero tranzitných uzlov

Problémom pri prenose cez viacero tranzitných uzlov je, že keď signály prijaté v cieľovom uzle sú zložené zo signálov aj od škodlivých aj kooperatívnych uzlov. Izolovanie týchto uzlov môže byť vykonávané pomocou detekcie maximálnej pravdepodobnosti, avšak nastávajú situácie keď je táto detekcia nespoľahlivá. Aj kooperujúce tranzitné uzly môžu byť považované za škodlivé, pretože signály prenášané skutočnými škodlivými uzlami môžu spôsobovať interferencie v cieľovom uzle.

Pri predpoklade neortogonálneho kooperatívneho prenosu cez prenosový kanál, signály preposielané cez tranzitné uzly sa na prijímači prekrývajú, čím je ťažšie identifikovať signál každého prenosu. Pokúsiť sa oddeliť signály je možné za pomoci detekcie maximálnej pravdepodobnosti a následne vyhodnotiť koreláciu signálov medzi každým tranzitným a cieľovýmuzlom. Predpokladajme však aj situáciu keď je zdrojový signál v cieľovom uzle veľmi slabý a tak namiesto výpočtu korelácie medzi zdrojom a symbolmi prichádzajúcimi z tranzitných uzlov vykonáme korelačný test so sekvenciou stopovacích symbolov známych ako priori (prednostné) v zdrojom aj cieľovom uzli. Stopovacie symboly sú generované generátorom pseudo-náhodných čísel (PRGN) založeným na určitom stopovacom kľúči, ktorý je známy zdrojovému aj cieľovému uzlu, ale skrytý pred tranzitnými uzlami.

Tieto symboly sú náhodne vkladané zdrojovým uzlom do dátového toku. Výpočtom korelácie stopovacích symbolov preposielaných každým tranzitným uzlom s ich odpovedajúcou správnou hodnotou, cieľový uzol dokáže identifikovať škodlivé uzly. Tento postup je znázornený diagrame na obr.3[12], [9].



Obr. 3 Proces detekcie škodlivých uzlov použitím stopovacích symbolov.

IV. ZÁVER

Tento článok predstavuje úvod do problematiky teórie hier ako aj bezpečnosti v sieťach, ktoré predstavujú evolúciu mobilných ad-hoc sietí. Pri poskytovaní bezpečnostných služieb je potrebné zabezpečiť ciele, ktoré sa vyžadujú pri prenose v každom type siete. Sú to dôvernosť (*confidentiality*), integrita (*integrity*), dostupnosť (*availability*), autentizácia (*authentication*), neodmietnuteľnosť (*non-repudation*), autorizácia (*authorization*) a dôveryhodnosť (trustworthiness). K naplneniu týchto požiadaviek je potrebné vytvárať bezpečnostné mechanizmy fungujúce nielen v typických MANET, ale aj v kooperatívnych sieťach. Tieto mechanizmy sú nevyhnutné na ochranu kooperujúcich sietí, ktoré vylučujú škodlivé a sebecké

uzly zo smerovania v sieti. Využitím základných algoritmov teórie hier môžeme účinne poskytovať mechanizmy ochrany dát pri prenose medzi zdrojovým a cieľovým uzlom.

POĎAKOVANIE

Táto práca bola vytvorená realizáciou projektu Rozvoj Centra informačných a komunikačných technológií pre znalostné systémy (kód ITMS projektu: 26220220230) na základe podpory operačného programu Výskum a vývoj, financovaného z Európskeho fondu regionálneho rozvoja (50 %) a projektom INDECT (FP7-No.218086) (50%).

LITERATÚRA

[1] A. Nosratinia, T. Hunter, A. Hedayat, „CooperativeCommunication in WirelessNetworks“,Universityof Texas, Dallas
[2] K.J.R. Liu, K. Sadek, A.S.Weifeng, A.Kwasinski, „CooperativeCommunications and Networking“ Cambridge University Press 2009, ISBN-13 978-0-511-46548-2
[3] A. E. Khandani, J. Abounadi, E.Modiano, L.Zheng, „CooperativeRouting in WirelessNetworks“. Laboratoryforinformation and decisionsystem Massachusetts InstitudeofTechnology
[4] L.Fulu, K. Wu,A. Lippman, „Energy-EfficientCooperativeRouting in Multi-hopWireless Ad Hoc Networks“,TheMediaLaboratory, MIT, Cambridge, MA USA, 2006
[5] P. Y.W. Hong, W.J. Huang, J. Wan Jen, C.C. Kuo: CooperativeSommunications and Networking. Technologies and systemdesign, 2010. ISBN 978-1-4419-7193-7
[6] M. Felegyhazi, J.P.Hubaux, „Gamy Theory in WirelessNetwork, a Tutorial“. EPFL, Switzerland, 2006
[7] M. Mehrandish., H. Otrok, M. Debbabi, C. Assi and P.Bhattacharya, “A Game TheoreticApproach to DetectNetworkIntrusions: TheCooperativeIntrudersScenario”, Proc. 49th annualof IEEE GLOBECOM, San Francisco, IEEE press, 2006.
[8] P. Michiardi and R. Molva, “AnalysisofCoalitionFormation and CooperationStrategies in Mobile Ad hoc Networks”, Journalof Ad hoc Networks, Elsevier,pp:193-219, 2005.
[9] Su W, Sadek A K, RayLiu K J. Cooperativecommunicationprotocols in wirelessnetworks: performanceanalysis and optimum powerallocation. WirelPersCommun, 2008, 44(2):181–217
[10] L. Buttyán, J.P. Hubaux, „Security and cooperation in wirelessnetworks“. Thwartingmalicious and selfishbehaviourinttheageofubiquitouscomputing. Version 1.5.1., 2007. ISBN 978052183710
[11] Z. Li, H. Shen, „AnalysistheCooperationStrategies in Mobile Ad hoc Networks“, Department ofComputerScience and ComputerEngineering, Universityof Arkansas, 2008
[12] P.Y.W Hong,W.J. Huang, C.C. Kuo, „CooperativeSommunications and Networking. Technologies and systemdesign“, 2010. ISBN 978-1-4419-7193-7
[13] C. Hausl , J. Hagenauer, „Relaycommunicationwithhierarchicalmodulation“, IEEE CommunicationsLetters, vol. 11, pp. 64-66, Jan., 2007
[14] R. Mahajan, M. Rodrig, D. Wetherall, and J. Zahorjan. “SustainingCooperation in Multi-HopWirelessNetworks.” In Proceedingsofthe 2nd Conference on Symposium on NetworkedSystemsDesignImplementation, Vol. 2, pp. 231-244, 2005.
[15] M. Kim, M. Medard, J. Barros, and R. Kotter, “Analgebraicwatchdogforwirelessnetworkcoding,” Proc. of IEEE InternationalSymposium on InformationTheory, June2009.

Three-Dimensional OFDM in Frequency-Selective Fading Channel

Ján PASTIRČÁK, Denis DUPÁK, Dušan KOCUR, Juraj GAZDA

Department of Electronics and Multimedia Communications, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Slovak Republic

pastircak.jan@tuke.sk, denis.dupak@tuke.sk, dusan.kocur@tuke.sk, juraj.gazda@tuke.sk

***Abstract*—Recently, MD base-band modulation schemes have been proposed as promising tools for the improvement of the power efficiency of optical and wireless communication systems. In this paper, we deal with 3D base-band modulation applied for the orthogonal frequency division multiplexing (OFDM) wireless transmission system. In the case of 3D modulation schemes, each signal symbol is expressed by three real coordinates. It results in, that 2D structure of subcarriers has been used for OFDM [1]. In the next, OFDM of that kind will be referred to as 3D+2D OFDM. On the other hand, the conventional OFDM system exploiting 2D base-band modulation scheme in the combination with 1D structure of subcarriers will be denoted as 2D+1D OFDM. The intention of this paper is to compare the basic properties of 2D+1D OFDM and 3D+2D OFDM modulation schemes if they are performed through AWGN and frequency selective time-invariant transmission channel. The obtained results expressed by the bit error rate (BER) indicate that 3D+2D OFDM overcomes clearly 2D+1D OFDM.**

***Keywords*— OFDM, 3D modulation, 2D Fast Fourier Transformation**

I. INTRODUCTION

The demands for high data rates transmission over multipath radio channels have increased rapidly. To fulfill this requirement, utilization of multi-carrier based transmission techniques seems to be an inevitable solution. Very promising modulation technique which satisfies the communication system requirements is OFDM. OFDM is spectrally efficient modulation technique that has recently gained much popularity due to high speed and very effective data transmission over multipath fading channels. OFDM based transmission systems are characterized by large number of benefits in comparison with traditional schemes. This makes it superior compared to previously introduced wireless standards [1].

On the other hand, OFDM has also some well known drawbacks. To perform interference free demodulation the subcarriers have to remain orthogonal, otherwise system performance will degrade. Therefore, OFDM is very sensitive, to frequency offset originating from the mismatch of transmitter and receiver local oscillators. The time domain OFDM signal is a sum of a large number of complex sinusoids, which means that, according to the central limit theorem, the amplitude distribution will be Gaussian, leading to a large peak to average power ratio (PAPR) of the signal. Hence, a power amplifier with a relatively large linear range is required, otherwise non-linear effects will severely degrade the system performance [2].

Up to this date, several strategies decreasing OFDM transmission system drawback impact and have been proposed. A new approach in OFDM systems is the application of 3D base-band modulations instead of conventional 2D modulation. 3D modulation schemes are characterized by the power efficiency due to increasing of the minimum Euclidean distance (MED) of the constellation diagram symbols. Thus it can be expected that the application of 3D base-band modulations for OFDM schemes can improved some performance properties of OFDM transmission systems. On the other hand, 3D+2D OFDM modulation schemes have to employ 2D structure of subcarriers and hence 2D IFFT and 2D FFT have to be implemented in the transmitter and receiver, respectively. Therefore, 3D+2D OFDM will be characterized with the higher computational complexity than that of 2D+1D OFDM.

In the literature, the several papers on 3D+2D OFDM have been only devoted. A representative performance of 3D+2D OFDM and 2D+1D OFDM has not been yet published. Therefore we would like to present the introductory study intent on an illustration of 3D+2D OFDM and 2D+1D OFDM performance in this paper. More precisely, we will try to compare

the properties of 2D+1D OFDM and 3D+2D OFDM modulation schemes if they are performed through AWGN and frequency selective time-invariant transmission channel. The obtained results expressed by the bit error rate (BER) will indicate that 3D+2D OFDM has the potential to overcome 2D+1D OFDM.

The rest of the paper is organized as follows. In the next section, the basic structure of 3D+2D OFDM is outlined. A frequency selective channel model applied in our simulations will be presented in Section 3. The pilot based channel estimation and the channel equalization are introduced in Section 4. Section 5 is devoted to the computer simulations intent on the illustration of the 3D+2D OFDM and 2D+1D OFDM performance. Finally, conclusions and final remarks to this contribution are drawn in Section 6.

II. 3D + 2D OFDM

Conventional OFDM system uses two-dimensional signal mapper and one-dimensional IFFT to create set of complex one-dimensional OFDM symbols. Reference [3] presents a new approach to the OFDM. Instead of using two-dimensional signal mapper, it uses three-dimensional mapper in combination with two-dimensional IFFT to create three-dimensional time domain OFDM symbol. We can represent constellation symbols of the M-ary 3D signal mapper as

$$S=\begin{bmatrix} x_0 & x_1 & \dots & x_{N-1} \\ y_0 & y_1 & \dots & y_{N-1} \\ z_0 & z_1 & \dots & z_{N-1} \end{bmatrix} \tag{1}$$

where x_i, y_i, z_i are coordinates of constellation symbols in 3D sapce. The data symbols enter to the 2D-IFFT block, where are modulated to the subcarriers frequencies. The representation of the 2D-IFFT operation can be expressed as

$$s(n_1,n_2)=IFFT(S)=\frac{1}{N_1N_2}\sum_{k_2=0}^{N_2-1}\sum_{k_1=0}^{N_1-1}S(k_1,k_2)\cdot\exp\left[j2\pi\left(\frac{n_2k_2}{N_2}+\frac{n_1k_1}{N_1}\right)\right] \tag{2}$$

where N_1 is number of subcarriers and N_2 is number of dimensions of signal mapper (N_2=3 for 3D signal mapper). The block scheme of 3D+2D OFDM transmission system is in Fig.1. Other applied blocks will be described in the section 3 and 4.

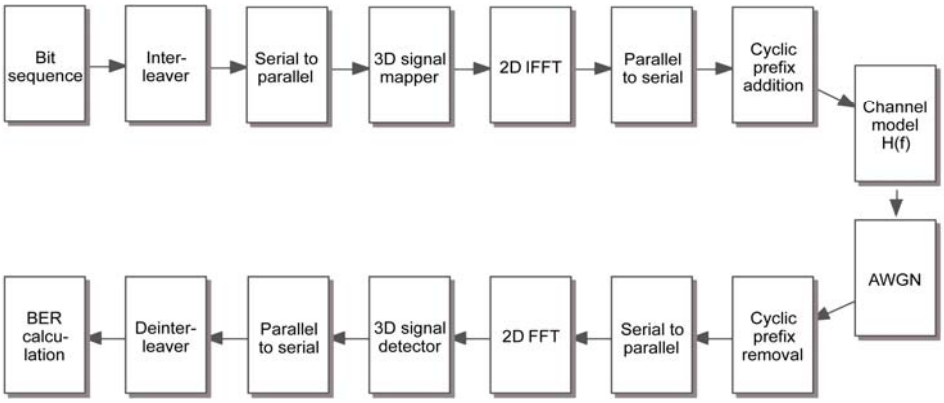


Fig 1.: 3D+2D OFDM transmission system

III. FREQUENCY-SELECTIVE CHANNEL MODEL

If the transmitted signal is propagated through the multipath channel, the received signal will contain signal components with various gains and delay. It can be expressed as

$$r(t)=A_0s(t-t_0)+A_1s(t-t_1)+\dots+A_ns(t-t_n) \tag{3}$$

where ***r(t)*** is the received signal, $A_0,A_1,\ldots,A_n$ are the signal components gains and $t_0,t_1,\ldots,t_n$ are the signal components time delays. In general A_k and t_k are variables that are changing in time but for simplification we can consider them as constants and then the channel can be modeled as a linear time-invariant filter. As we sample the received signal in the discrete time points, than the channel model will be a FIR filter with the impulse response

$$h[n] = \sum_{k=0}^{N} a_k \delta[n-k] \,, \tag{4}$$

where N is the length of channel impulse response, a_i are the filter coefficients and $\delta[n]$ is the Kronecker delta function. Let s_T is the OFDM symbol, which can be represented as a $3\times N$ matrix, than the received OFDM symbol s_R can be expressed as

$$s_R[i,n] = h[i,n] * s_T[i,n] \tag{5}$$

where i is the index of the signal dimension (for 3D+2D OFDM $i=\{1,2,3\}$). Then the i-th dimension of the received symbol is expressed as a convolution between i-th dimension of the transmitted symbol and i-th channel impulse response.

The channel frequency response is the Fourier Transformation of the channel impulse response

$$H[i,f] = F\{h[i,n]\} \tag{6}$$

Then the received OFDM symbol in the frequency domain can be expressed as

$$S_R[i,f] = H[i,f] \circ S_T[i,f] \tag{7}$$

where operator "$\circ$" represents Hadamard multiplication (multiplication between individual components of the matrix) and f represents a subcarrier number. This model was used in the simulations of the 3D+2D OFDM in the frequency-selective channel, which were done for the purposes of this paper. The characteristics of frequency selective channel responce, which was used in simulations is in Fig. 2 [4].

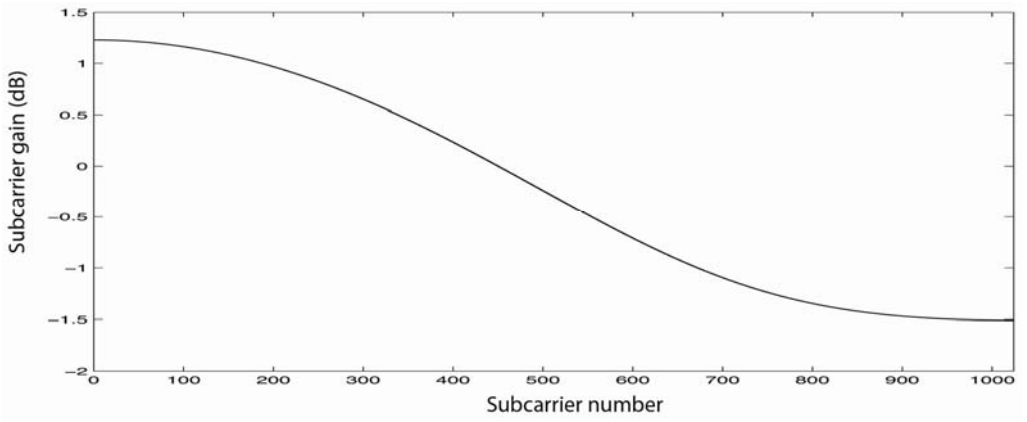


Fig 2.: Channel frequency response (N=1024)

IV. PILOT-BASED CHANNEL ESTIMATION AND EQUALIZATION

The pilot symbols are special data, which are known (value, localization in the time and frequency coordinates) both by the transmitter and the receiver. They are used for the channel estimation, transmitter synchronization, etc. If the total OFDM symbol is transmitted as a pilot symbol, the receiver can estimate the frequency response on the channel corresponding to the subcarrier frequency by (7) using Zero-forcing method.

If the receiver know the channel frequency response, it can make a frequency-domain equalization to compensate the distortion of the frequency-selective channel.

$$S_{\mathrm{R}_{eq}}[i,f] = \frac{S_R[i,f]}{H_e[i,f]} \tag{8}$$

where $\boldsymbol{S_{Req}}$ represents an equalized received symbol. Then the error of the equalization can be expressed as

$$e_{eq} = \frac{1}{3 \cdot N} \sum_{k_1=1}^{3} \sum_{k_2=1}^{N} \{S_{\mathrm{R}_{eq}}[k_1, k_2] - S_T[k_1, k_2]\}^2 \tag{9}$$

and it depends on the E_b/N_0 [5].

V. SIMULATION RESULTS

A. AWGN channel simulation

To evaluate the error performance of the OFDM system with 3D signal mapper and 2D IFFT we use simulations with 16-ary and 64-ary signal mapper and 1024 subcarriers. In the results of the first simulation we can see relation between BER and E_b/N_0 for 2D+1D and 3D+2D OFDM transmission systems in the AWGN channel.

As we can see in Fig.3. 3D+2D OFDM system overperforms conventional 2D+1D OFDM systems. 3D-64 base band modulation provides performance gain about 8dB for BER=10^{-4} in comparison with 64-QAM modulation. Performance gain for 3D-16 modulation is about 5dB for BER=10^{-4} in comparison with 16-QAM.

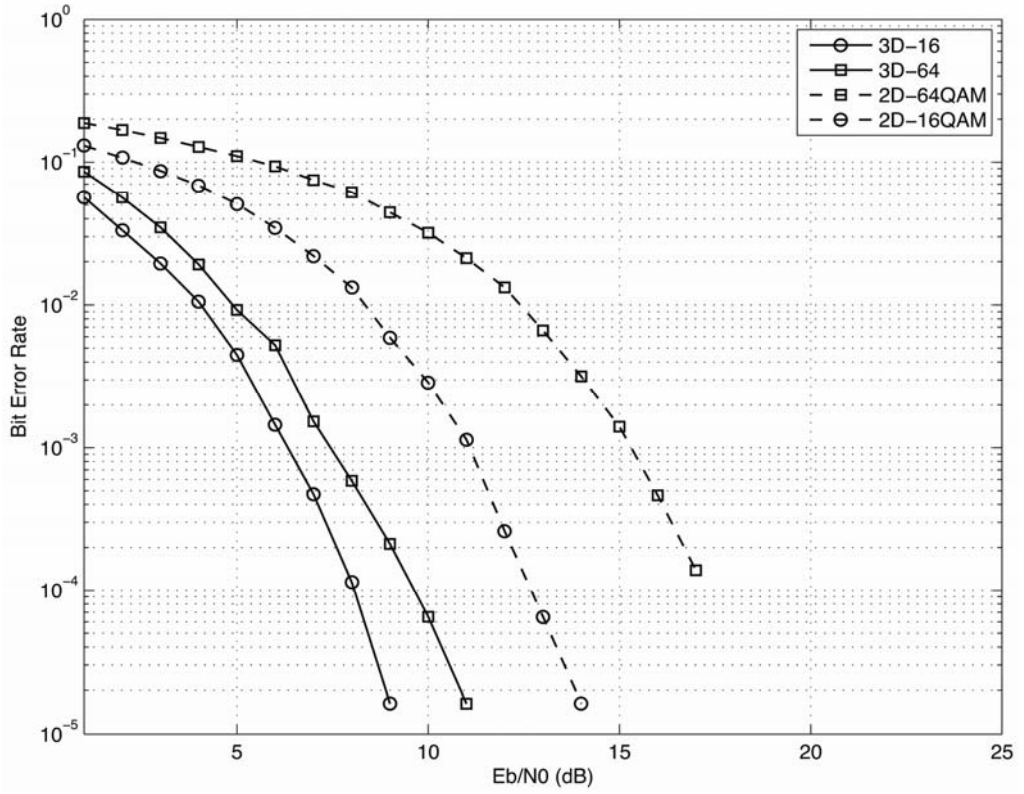


Fig 3. Simulation results I. BER vs. Eb/N0 for AWGN channel

B. Frequency-selective channel simulation

In the frequency-selective channel simulation, the one-tap frequency domain equalizer based on the pilot symbol channel response estimation described in [6] is used for 3D+2D OFDM. In order to employ this equalization method and to avoid intercarrier interference (ICI), the cyclic prefix insertion has been implemented. The cyclic prefix insertion is outlined in Fig. 4, where data transmitted within this guard time are simply a copy of the end of the OFDM block. Data transmitted in this guard time are simply a copy of the end of the OFDM block and it is called cyclic prefix. Cyclic prefix also ensures that convolution between channel impulse response and transmitted signal will have a form of the circular convolution and this is equivalent to

multiplication in frequency domain. This processing allows the equalizer to do the inverse channel filtering in the frequency domain as a simple multiplication [6].

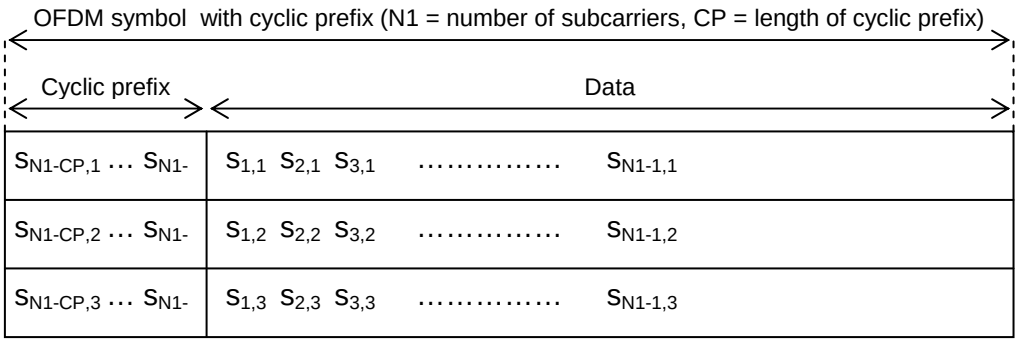


Fig 4. 3D+2D OFDM symbols with cyclic prefix

The simulation parameters are $N = 1024$ (number of subcarriers) and length of the cyclic prefix is $L_{CP} = 128$. The channel response used in this simulation is shown in Fig.2.

As we can see in Fig.5, the effect of equalization is very noticable. Received symbols modulated with 3D modulation are without equalization almost undetectable. Equalization eliminates ICI interference and increases BER performance of 3D+2D OFDM transmission system. BER performance comparison for 3D-16-ary and 3D-64-ary modulations with and without equalization is in Fig.5.

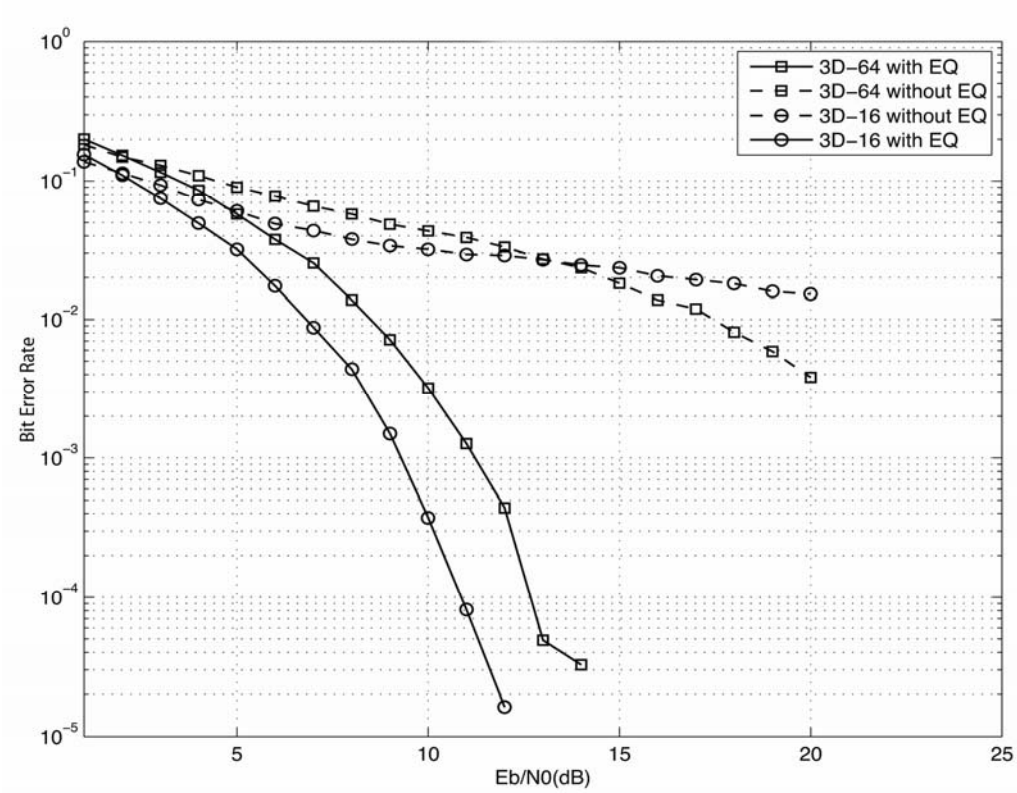


Fig 5. Effect of the frequency domain equalization for 3D+2D OFDM

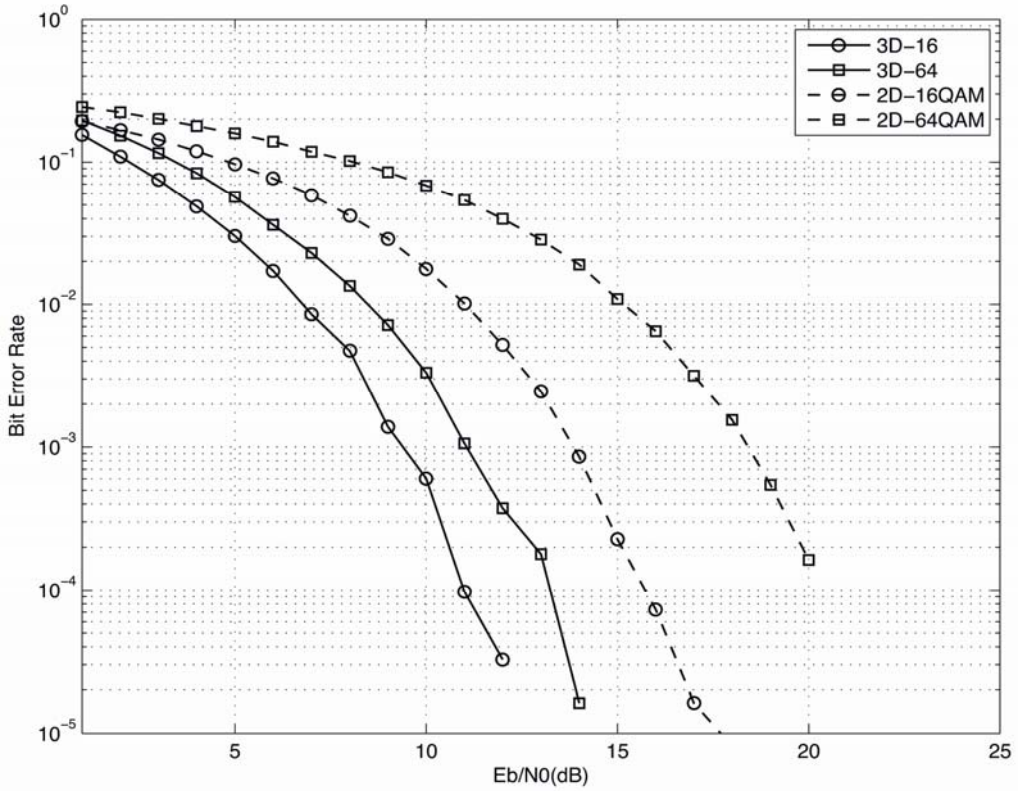


Fig 6. Simulation results II. BER vs. Eb/N0 for the frequency-selective channel

Figure 6 shows the comparison of BER performance 3D+2D OFDM and 2D+1D OFDM for the frequency selective channel. As it can be seen from this figure, 3D+2D OFDM transmission system outperforms conventional OFDM systems. Here, 3D-64 base band modulation provides performance gain about 7dB for BER=10^{-4} in comparison with 64-QAM modulation. The performance gain for 3D-16 modulation is about 5dB for BER=10^{-4} in comparison with 16-QAM.

VI. Conclusion

3D+2D OFDM is the novel modulation technique possessing better BER performance than the conventional OFDM in both AWGN and frequency-selective channel. This property is due to the higher MED of 3D signal constellation schemes. The better performance of 3D+2D OFDM is reached at the cost of its higher complexity in comparison with that of 2D+1D OFDM schemes. The performance of 3D+2D OFDM indicates very clearly, the MD modulation schemes have the great potential to be applied for the efficiency improvement of the conventional OFDM scheme.

References

[1] E.Dahlman, S.Parkvall, "*3G Evolution, HSPA and LTE for Mobile Broadband. Academic Press*", Oxford UK, 2007.
[2] H.G. Myung, D. J. Goodman, "Single Carrier FDMA: A New Interface for Long Term Evolution", John Wiley & Sons, Ltd., 2008.
[3] S.G. Kang, "An OFDM with 3-D Signal Mapper and 2-D IDFT Modulator"*, IEEE Communication Letter, vol 12., no 12. December 2009, pp. 871-873*
[4] J. van de Beek, O. Edfors, M. Sandell, S. K. Wilson, P. O. Börjesson, "On Channel Estimation in OFDM systems", *Proceedings of Vehicular Technology Conference*, vol. 2, Chicago, USA, September 1995, pp. 815-819
[5] M. S. Akram, "Pilot-based Channel Estimation in OFDM Systems", Master Thesis, Nokia Mobile Phones, 2007
[6] E. Ch. Choi, J. D. Huh, S. G. Kang, "A New Design of OFDM for Ultra-Wideband Systems", *Consumer Electronics, 2009. ICCE '09. Digest of Technical Papers International Conference on* , vol., no., pp.1-2, 10-14 Jan. 2009.

Farebné filtre v predspracovaní obrazu pre optický korelátor

František RAKOCI, Ján TURÁN, Ľuboš OVSENÍK

Katedra elektroniky a multimediálnych telekomunikácií, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

ferorak@gmail.com, jan.turan@tuke.sk, lubos.ovsenik@tuke.sk

***Abstrakt* — Tento článok popisuje aplikáciu farebných filtrov v predspracovaní obrazovej scény pre vstup optického korelátora. Optický korelátor dokáže pracovať iba s binárnymi vstupnými obrazmi. Farebné filtre zabezpečujú segmentáciu obrazu podľa jednej definovanej farby, alebo podľa farieb z určitého rozsahu. Následným prahovaním je vytvorený binárny obraz. Ten pozostáva z geometrických tvarov, ktoré zodpovedajú tvarom sledovaných objektov, napr. dopravných značiek. Štatistický rozptyl farieb zo snímanej scény oproti ideálnej farbe je možné znížiť konverziou 3 rozmerného RGB modelu do dvojrozmerného normalizovaného modelu farieb. V článku je opísaný postup spracovania vstupnej scény do binárneho obrazu a zníženie štatistického rozptylu farieb. Taktiež sú uvedené softvérové knižnice pre prácu s digitálnym obrazom, ktoré obsahujú funkcie pre spracovanie.**

***Kľúčové slová* — farebný filter, segmentácia, štatistický rozptyl, prahovanie, optický korelátor**

I. ÚVOD

Softvérové farebné filtre sú hlavnou súčasťou systémov počítačového videnia [2]. Sú využívané v robotike, asistenčných systémoch, medicíne a pod. Farebné filtre slúžia na analýzu objektov získaných pomocou digitálnej kamery. Na základe zvoleného kritéria dokážeme snímok upraviť, zvýrazniť jeho časti, prípadne vyselektovať objekty záujmu. V našom výskume sa zaoberáme rozpoznávaním dopravných značiek s využitím optického korelátora [3]. Optický korelátor dokáže pomocou sústavy optických šošoviek v reálnom čase realizovať Fourierovu optickú transformáciu obrazu a určiť mieru podobnosti dvoch vstupných objektov.

V článku prezentujeme efektívny spôsob segmentácie obrazu na základe farby. Táto segmentácia zabezpečí prevod obrazu do binárneho tvaru so zobrazením kontúry rozpoznávaných objektov. V cykle segmentácie sa zabezpečí odstránenie zhlukov farieb a určenie polohy rozpoznávaného objektu v snímke.

V prvej časti článku sú opísané farebné modely, ktorými možno reprezentovať farbu každého obrazového prvku (o.p.). Ďalšia časť je venovaná zníženiu štatistického rozptylu konverziou RGB farieb do normalizovaného modelu. Potom je uvedený postup segmentácie podľa zvolenej farby. Na záver sú uvedené a stručne charakterizované softvérové knižnice pre prácu s digitálnym obrazom.

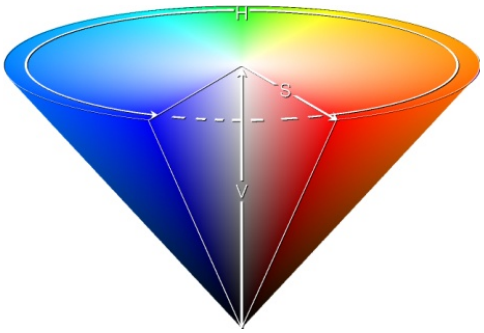
II. ZÁKLADNÉ FAREBNÉ MODELY

Digitálne obrazy predstavujú pole o.p. (maticu) zaznamenaných v počítačovej pamäti. Tieto sú následne zobrazené na obrazovke. Každý o.p. je charakteristický svojou polohou v tomto poli a svojou farbou. Vo všeobecnosti, farba je elektromagnetické žiarenie s určitou vlnovou dĺžkou. Zmenou hodnoty tejto vlnovej dĺžky sa mení aj vnímanie tejto farby ľudským okom. Farebné modely opisujú základné farby a schémy ich miešania do výslednej farby.

A. HSV (HSB) model

HSV (Hue Saturation Value (Brightness)) model je model, ktorý najviac zodpovedá vnímaniu farieb ľudským okom. Farby v tomto modeli sú reprezentované tromi hodnotami. Prvou hodnotou je farebný tón, alebo odtieň (H - Hue). Môže sa pohybovať v rozsahu od 0 do 360°. 0° v tomto prípade označuje červenú farbu. Ostatné farby sú dané usporiadaním na štandardnom farebnom kruhu. Druhou hodnotou, ktorá popisuje farbu v HSV modeli je nasýtenie (S - Saturation). Táto hodnota udáva množstvo šedej farby v pomere k základnému

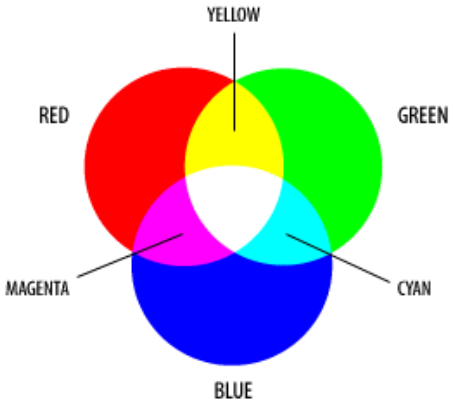
odtieňu. Udáva sa v hodnotách od 0 % do 100 %. Na farebnom kruhu nasýtenie stúpa od stredu ku okraju. Poslednou hodnotou je hodnota jasu (V – Value, alebo B - Brightness), teda množstvo bieleho svetla. Hodnota jasu vyjadruje, koľko svetla daná farba odráža [6].



Obr. 1 Znázornenie farebného modelu HSV [6].

B. RGB model

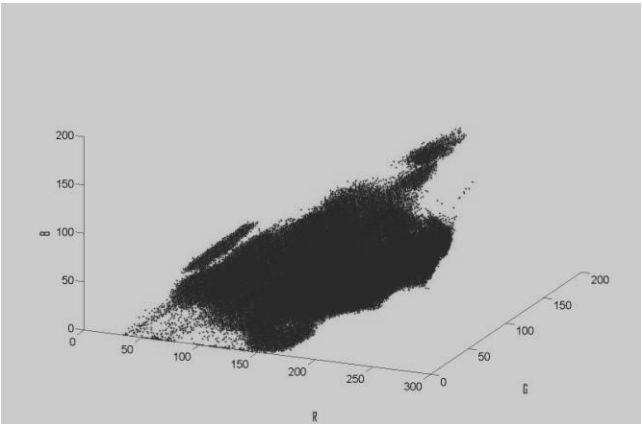
Interpretácia farieb v počítačovom spracovaní je najčastejšie realizovaná kombináciou troch základných farieb. Každá hodnota reprezentuje podiel príslušnej základnej farby R (Red) – červená, G (Green) – zelená, B (Blue) – modrá. Kombináciou základných farieb vzniká výsledná farba. Keďže každý bajt môže kódovať hodnoty od 0 do 255, znamená to, že kombináciou troch základných farieb s rôznymi hodnotami môžeme dosiahnuť až 16777216 rôznych výsledných farieb [5].



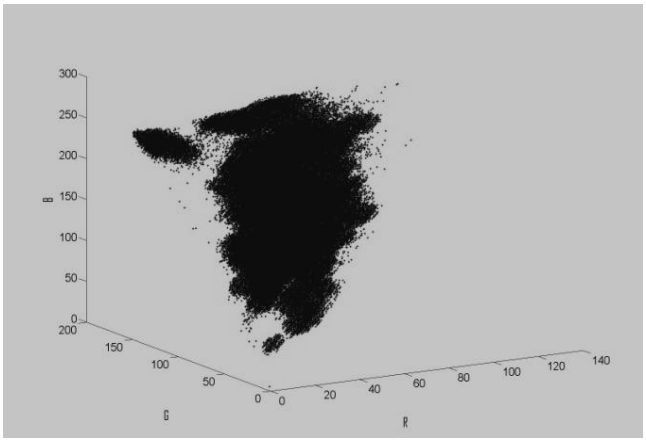
Obr. 2 Znázornenie farebného modelu RGB [5].

III. NORMALIZÁCIA ZÁKLADNÝCH FARIEB

Reprezentácia farby objektov, ktoré sú zaznamenané kamerou môže byť odlišná v závislosti od rôznych podmienok. Tento jav je spôsobený vplyvom počasia, v dôsledku rôznych parametrov kamier, alebo uhlom dopadajúceho svetla.



Obr. 3 Rozptyl troch miliónov červených vzoriek [1].



Obr. 4 Rozptyl troch miliónov modrých vzoriek [1].

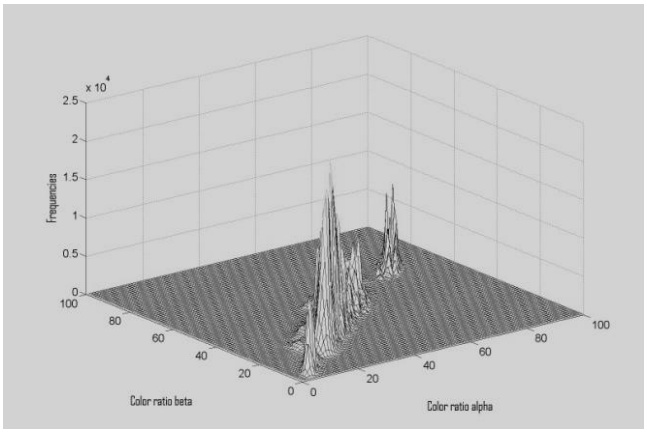
Na Obr. 3 a Obr. 4 je znázornený rozptyl vo farebnom priestore RGB (Red Green Blue). Opisujú rozptyl troch miliónov vzoriek dopravný značiek s červenou a modrou farbou.

RGB farebnú vzorku môžeme transformovať do dvojrozmerného modelu. Dva farebné kanály znormalizujeme podľa tretieho dominantného kanálu nasledujúcim spôsobom [1]:

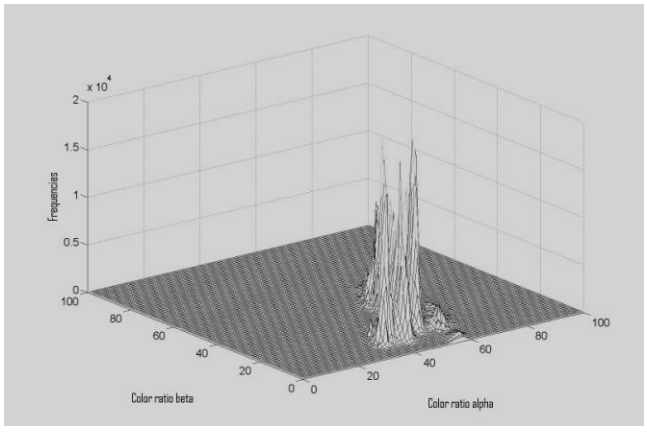
$$\text{Červená farba:} \quad \alpha = \frac{g}{r}; \qquad \beta = \frac{b}{r}; \qquad (1)$$

$$\text{Modrá farba:} \quad \alpha = \frac{r}{b}; \qquad \beta = \frac{g}{b}; \qquad (2)$$

Táto normalizácia výrazne redukuje rozptyl farebnosti spôsobený prostredím a parametrami kamery. Zredukovaním farebného priestoru do dvoch rozmerov (2D) sa okrem jednoznačnejšej farebnej interpretácie výrazne znížia aj požiadavky na výpočtovú kapacitu. Výsledný farebný rozptyl 2D reprezentácie farieb je znázornený na Obr. 5 a Obr. 6.



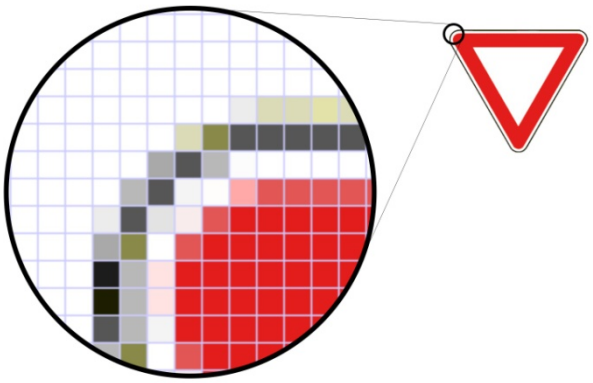
Obr. 5 Rozptyl vzoriek červenej farby v normalizovanom priestore [1].



Obr. 6 Rozptyl vzoriek modrej farby v normalizovanom priestore [1].

IV. SEGMENTÁCIA VSTUPNEJ SCÉNY

Našim cieľom je vybrať zo scény objekty nášho záujmu. Charakteristický tvar týchto objektov vo forme binárneho obrazu následne poskytneme optickému korelátoru ako vstup. Každý objekt na scéne je zobrazený pomocou určitého zoskupenia o.p. Každý o.p. je charakteristický svojou polohou a farbou.



Obr. 7 Detail dopravnej značky so zobrazením o.p..

O.p. zobrazujúce jeden objekt majú niektoré spoločné znaky. Prvým spoločným znakom týchto o.p. je, že sa nachádzajú navzájom v bezprostrednej vzdialenosti. Druhým znakom je ich podobnosť vo farbe.

V rámci jednej scény sa môže vyskytnúť viacero objektov, spĺňajúcich zvolené parametre. Pre spoľahlivosť spracovania budeme brať do úvahy vždy iba najrozľahlejší objekt na scéne.

V tomto kroku nastáva ďalší problém. Potrebujeme vyčleniť najväčší vyhovujúci objekt zo scény. Je preto potrebné odlíšiť o.p. zobrazujúce jeden objekt od o.p. zobrazujúcich ďalšie objekty.

A. Voľba farebného kritéria

Prvou fázou segmentácie je zvolenie vhodných farebných kritérií pre hľadanú dopravnú značku. V prípade normalizovaného modelu toto kritérium predstavuje štvorec v 2D súradnicovej sústave. Tento štvorec vymedzuje minimálne a maximálne hodnoty pre každú súradnicu. V prípade, že je potrebné pracovať s farbami mimo normalizovaných základných farieb, možno zvoliť iný model, ktorý bude vyhovovať konkrétnemu účelu.

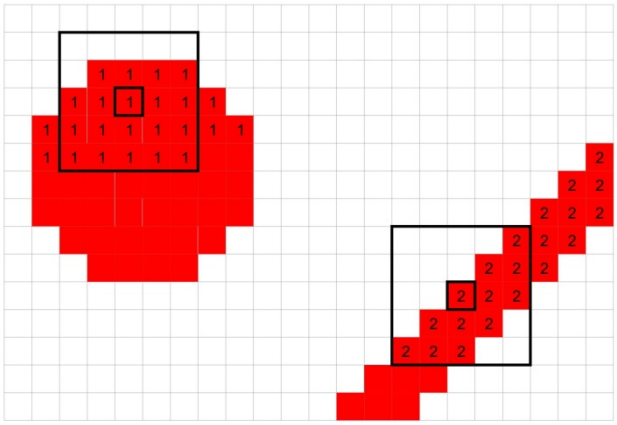
B. Indexovanie spojitých oblastí v obraze

Počiatočný bod indexovania je v ľavom hornom rohu obrazu. Postup indexovania spočíva v kontrole farebného kritéria pre každý pixel a jeho okolie.

Môže nastať niekoľko prípadov:

- Aktuálny o.p. nespadá do zadaného farebného kritéria. V tomto prípade je tento o.p. ignorovaný a pokračuje sa v kontrole ďalších o.p.
- Aktuálny o.p. spadá do zadaného kritéria, ale v jeho okolí sa nenachádza žiaden iný o.p. s podobnou farbou. Tento o.p. je taktiež ignorovaný, pretože sám o sebe nepredstavuje žiaden zmysluplný objekt.
- Aktuálny o.p. a jeho susedné o.p. spadajú do zadaného farebného kritéria. Susedné o.p. sú navyše označené identifikátorom. V tomto prípade je aktuálny o.p. označený tým istým identifikátorom, ako jeho susedné o.p.
- Aktuálny o.p. a jeho susedné o.p. spadajú do zadaného kritéria. Žiaden susedný o.p. však ešte nemá svoj identifikátor. Vtedy sa všetkým aktuálnym o.p. priradí identifikátor ďalšej oblasti.

Vďaka takto označeným oblastiam dokážeme vytvoriť binárny obraz ľubovoľnej oblasti v scéne. Ten potom môže slúžiť ako plnohodnotný vstup pre optický korelátor [3].



Obr. 8 Indexovanie spojitých oblastí.

V. SOFTVÉROVÉ NÁSTROJE

A. OpenCV

Pre spracovanie obrazu možno využiť knižnicu Open CV [4]. Open CV je knižnica programových funkcií, ktoré sú zamerané hlavne na počítačové videnie v reálnom čase. Open CV bola pôvodne napísaná v jazyku C, no v súčasnosti má plné rozhranie C++ a ďalší vývoj je realizovaný taktiež v tomto jazyku. V knižnici je implementované aj rozhranie pre jazyk Python. Open CV je vhodná pre vývoj množstva typov grafických aplikácií, napr. segmentácia a rozpoznávanie, sledovanie pohybu, multikamerová kalibrácia, atď.

Open CV má modulárnu štruktúru, čo znamená, že celý balík obsahuje viacero zdieľaných, alebo statických knižníc [4].

Knižnica ImgProc

ImgProc je modul Open CV pre spracovanie obrazu, ktorý zahŕňa lineárne a nelineárne filtrovanie obrazu, geometrické obrazové transformácie, konverzie medzi farebnými priestormi, spracovanie histogramov, atď. Táto knižnica obsahuje funkcie využiteľné pre farebnú filtráciu obrazov. Napríklad CvtColor() pre konverziu medzi farebnými priestormi, alebo funkcia Threshold() pre aplikáciu jednoúrovňového prahovania vybraného farebného kanálu [4].

B. AForge.NET

AForge.NET je ďalšou knižnicou, ktorá je využiteľná pre náš účel. Je to voľne širiteľná C# knižnica navrhnutá pre vývojárov a výskumníkov na poli počítačového videnia a umelej inteligencie – spracovanie obrazov, neurónové siete, genetické algoritmy, fuzzy logika, strojové učenie, robotika, atď.

Tento framework sa skladá zo skupiny knižníc a ukážkových aplikácií, ktoré demonštrujú ich možnosti. [8]

AForge.Imaging

AForge.Imaging je najväčšou knižnicou z celého frameworku. Obsahuje rôzne procedúry pre spracovanie obrazu, ktoré sú zamerané na prácu ako oprava poškodeného obrazu, alebo počítačové videnie. Obsahuje taktiež veľmi užitočné funkcie pre náš účel, ktoré dokážu filtrovať pixely podľa farby.

AForge.Video

Táto knižnica obsahuje rôzne knižnice, ktoré poskytujú prístup k video sekvenciám. Dokážu napríklad zabezpečiť prístup k USB kamere, zápis a čítanie video súborov, atď.[8].

VI. ZÁVER

Farebné filtre sú stále dôležitejšou súčasťou počítačového videnia. Cieľom výskumu v oblasti farebných filtrov je čím viac sa priblížiť k princípu vnímania objektov ľudským okom. Počítačové videnie pomocou farebných filtrov by mohlo byť doplnené o rozpoznávanie objektov na základe pohybu objektu vzhľadom na pozadie scény. Vhodnou voľbou farebných modelov možno dosiahnuť menší rozptyl vzoriek z obrazov dopravných značiek. Systém s najlepšími výsledkami možno využiť v systéme inventarizácie dopravných značiek. Ten bude schopný automaticky rozoznať typ dopravnej značky, vyhodnotiť jej stav a zapísať dostupné informácie o tejto značke do databázy.

POĎAKOVANIE

Podporujeme výskumné aktivity na Slovensku/ Projekt je spolufinancovaný zo zdrojov EÚ. Tento článok bol vypracovaný v rámci projektu "Centrum excelentnosti integrovaného výskumu a využitia progresívnych materiálov a technológií v oblasti automobilovej elektroniky", ITMS 26220120055 (50%). Táto práca bola taktiež podporovaná a realizovaná z projektu Rozvoj Centra informačných a komunikačných technológií pre znalostné systémy (ITMS kód: 26220120030) na základe podpory operačného programu Výskum a vývoj financovaného z Európskeho fondu regionálneho rozvoja (50%).

LITERATÚRA

[1] Z. Hu, "*Intelligent Road Sign Inventory (IRSI) with Image Recognition and Attribute Computation from Video Log*", Kyoto, Japan, 2011.
[2] M. Petruš, "*Optical correlator at video driver assistance system*", Diploma work, Technical University of Košice, Košice, Slovakia, 2011.
[3] Q. Ye, W. Gao, W. Zeng, "Color Image Segmentation Using Density-Based Clustering", *ICASSP 2003*, China, 2003.
[4] *Open CV library*, published on: http://opencv.willowgarage.com/wiki/
[5] *The RGB (CMY) color model*, published on: http://dba.med.sc.edu/price/irf/Adobe_tg/models/rgbcmy.html
[6] *HSV model*, published on http://cs.wikipedia.org/wiki/HSV
[7] *Farebný model*, published on: http://sk.wikipedia.org/wiki/Farebn%C3%BD_model
[8] *AForge.NET*, published on: http://www.aforgenet.com/framework/

Inventarizačný systém dopravných značiek s využitím optického korelátora

František RAKOCI, Ľuboš OVSENÍK, Ján TURÁN

Katedra Elektroniky a multimediálnych telekomunikácií, Fakulta Elektrotechniky a Informatiky, Technická Univerzita v Košiciach, Slovenská Republika

ferorak@gmail.com, lubos.ovsenik@tuke.sk, jan.turan@tuke.sk

***Abstrakt* — Tento článok opisuje použitie optického korelátora v inventarizačnom systéme dopravných značiek. Hlavná úloha navrhovaného systému je rozpoznávanie dopravných značiek a ich parametrov v zaznamenanom videu. Toto rozpoznávanie sa vykonáva na základe pripravenej databázy s využitím optického korelátora. Cieľom je navrhnúť zariadenie na automatickú inventarizáciu dopravných značiek. Zariadenie musí byť schopné zaznamenať existenciu značky vo video zázname a špecifikovať správnu polohu značky, ako aj jej stav.**

***Kľúčové slová* — Fourierova optická transformácia, dopravné značky, rozpoznávanie, tvar, farba**

I. Úvod

Dopravné značky sú veľmi dôležitou súčasťou dopravnej infraštruktúry. Ich úlohou je riadiť dopravu, navigovať vodičov do žiadanej lokality a zvyšovať ich bezpečnosť [1,3]. Dopravné spoločnosti musia pravidelne kontrolovať stav všetkých dopravných značiek, na každom úseku cesty, ako súčasť svojich služieb. Následne musia uskutočniť nevyhnutné opravy alebo výmenu poškodených dopravných značiek. Funkciou navrhovaného systému je vyriešiť tento problém automatickou inventarizáciou dopravných značiek. Každá dopravná značka je charakteristická svojim vlastným tvarom a reprezentáciou jej farebných segmentov. Špecifikované sú taktiež rozmery značiek, ale tie môžu byť navzájom rozdielne, v závislosti od typu cesty, na ktorej sú značky nainštalované. Druhou dôležitou vlastnosťou dopravnej značky je jej poloha. Tá môže byť relatívna, od určitého základného bodu, alebo to môže byť absolútna poloha daná súradnicami GPS. Možno rozoznať existenciu dopravnej značky v zosnímanej scéne. To možno vykonať porovnaním geometrických a kolorimetrických vlastností reálnej značky s vlastnosťami ideálnych dopravných značiek. Je taktiež možné efektívne zistiť stav tejto dopravnej značky podľa parametrov získaných z video záznamu. Použitím GPS zariadenia a laserového diaľkomera sa dá zabezpečiť určenie presnej polohy dopravnej značky [6].

Dôležitou súčasťou navrhovaného systému je Cambridgský optický korelátor. Cambridgský optický korelátor je schopný vyhodnotiť podobnosť dvoch obrazov. Korelátor používa Fourierovu optickú transformáciu a autokoreláciu obrazov zložených z ideálnej dopravnej značky a zosnímanej scény [7].

Pre optický korelátor je nevyhnutné predspracovanie obrazov, s ktorými sa bude pracovať. Výsledkom tohto predspracovania sú binárne obrazy s geometrickými tvarmi. Predspracovanie je urobené príslušným farebným filtrom a procesom segmentácie. Korelátor je schopný porovnať a určiť mieru podobnosti sledovaných obrazov či videí.

Stav dopravnej značky je možné určiť použitím geometrických vlastností značky z video záznamu. Vieme, že video záznam je zložený zo sekvencie viacerých obrazov. Na základe štatistiky medzi týmito obrazmi možno zvýšiť pravdepodobnosť správneho rozoznania značky.

V prvej časti článku je opísaná funkcia jadra systému (optického korelátora). V druhej časti je opísané predspracovanie videa pre vstup optického korelátora. Na koniec je prezentovaný postup výpočtu polohy a sklonu dopravnej značky. V závere sú prezentované výhody a nevýhody diskutovaného systému.

ISBN 978-80-553-0890-6 © 2012 FEI TUKE

II. OPTICKÝ KORELÁTOR

Optický korelátor je zariadenie, ktoré je schopné porovnať dva vstupné obrazy a určiť ich podobnosť. Optický korelátor realizuje Fourierovu transformáciu optickou cestou [1].

Hlavné funkčné časti korelátora sú [7]:

A. Sústava optických šošoviek

Optické šošovky predstavujú najdôležitejšiu časť optického korelátora. Vykonávajú samotnú Fourierovu optickú transformáciu. Na začiatku je priestorovým optickým modulátorom vytvorený obraz. Obraz prechádzajúci cez šošovku vytvára Fourierovu optickú transformáciu na CMOS senzore.

B. Priestorový optický modulátor

Je to zariadenie, ktoré moduluje svetelný lúč na základe aktuálnej informácie vo vstupnom signáli. Svetelné lúče smerujú na povrch priestorového optického modulátora. Tekuté kryštály v každom obrazovom prvku (o.p.) modulátora sú riadené video signálom prostredníctvom DVI, alebo VGA rozhrania. Riadiaca jednotka adresuje každý o.p. a nastavuje optické parametre tekutých kryštálov. Toto riadenie je závislé od informácie zakódovanej vo video signáli. Priestorový optický modulátor je umiestnený v prednej ohniskovej vzdialenosti šošovky.

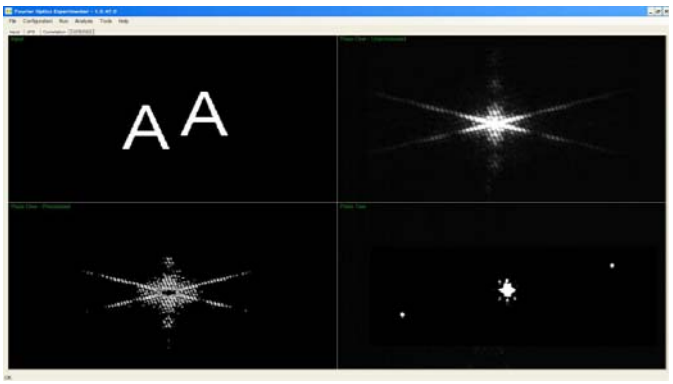
C. CMOS senzor

Je použitý na snímanie výsledku optickej korelácie. Po spracovaní sústavou šošoviek sú optické lúče nasmerované na plochu CMOS senzora. Tento výsledný obraz možno z CMOS senzora získať prostredníctvom Ethernet rozhrania. Prvou podmienkou pre jeho správnu funkciu je prepojenie medzi korelátorom a PC. Druhou podmienkou je správne nastavenie TCP/IP parametrov. CMOS senzor je umiestnený v zadnej ohniskovej vzdialenosti optickej šošovky.

D. Zdroj laserového žiarenia

Ako zdroj laserového žiarenia je použitá laserová dióda (LD), ktorá vyžaruje koherentné svetlo. Svetelný lúč je k Fourierovej optickej sústave privedený optickým vláknom. Laserové žiarenie je v optickom korelátore použité ako nosič informácie. Žiarenie je modulované konfiguráciou priestorového optického modulátora.

Na Obr. 1 je znázornený výstup optického korelátora vytvorený autokoreláciou vstupnej scény. Vstupná scéna je vytvorená z referenčnej vzorky a zosnímanej vzorky. Po autokorelácii je zobrazená najdominantnejšia jednosmerná zložka, prípadne jej postranné zložky. Postranné zložky reprezentujú zhodu, alebo podobnosť dvoch častí vstupnej scény. Použitím postranných zložiek možno určiť polohu zhody, alebo podobnosti na zosnímanej scéne oproti referenčnej vzorke.

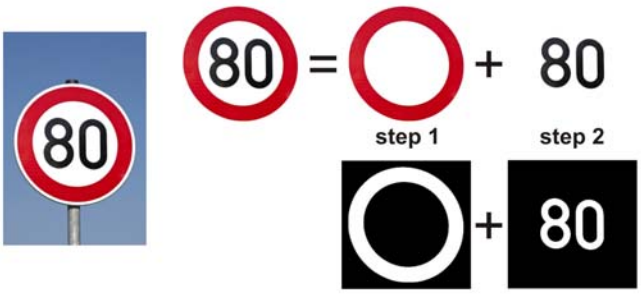


Obr. 1. Spracovanie vstupnej scény optickým korelátorom [1].

III. PREDSPRACOVANIE OBRAZU NA VSTUPE OPTICKÉHO KORELÁTORA

Pretože korelátor pracuje za pomoci koherentného laserového svetla, je schopný pracovať s binárnymi typmi obrazov. Preto musíme prispôsobiť referenčnú databázu a zosnímanú scénu. Toto prispôsobenie závisí od charakteristiky objektov, ktoré majú byť rozoznané. Ako príklad uvedieme dopravnú značku pre obmedzenie rýchlosti. V tomto prípade je nevyhnutné rozdeliť

rozpoznávanie a klasifikáciu dopravnej značky do dvoch krokov. V prvom rade je nevyhnutné vyhodnotiť existenciu dopravnej značky v zaznamenanej scéne. Toto vyhodnotenie môže byť zrealizované porovnaním scény s databázou možných geometrických tvarov. Tieto tvary reprezentujú kontúru danej dopravnej značky. V prípade značiek pre obmedzenie rýchlosti v Slovenskej republike je nevyhnutné rozoznať oblasť s tvarom kruhu. Po tejto operácii možno vykonať detailnejšiu klasifikáciu textu vo vnútri značky (Obr. 2).



Obr. 2. Segmentácia scény a prahovanie.

Pre efektívne spracovanie a pre určenie miery podobnosti musíme odstrániť zbytočné časti zo scény. V ideálnom prípade, výsledok tejto operácie by mala byť kontúra najväčšieho objektu, ktorý zodpovedá určeným farebným kritériám. Taktiež to môže byť celá scéna s podobnými objektmi. V prvom kroku musíme skonvertovať farebnú informáciu z každého o.p. z RGB do HSV modelu. Farby v RGB modeli sú definované kombináciou troch základných farieb pre každý o.p. Potom je výsledná farba vytvorená časťou červenej, zelenej a modrej farby. HSV model definuje farby podľa ich odtieňa, nasýtenia a intenzity. Použitím HSV modelu možno manipulovať s farbou v dimenziách čiastkových odtieňov. Nevyhnutnou modifikáciou, ktorú musíme vykonať je zníženie nasýtenia farebných odtieňov, ktoré nespadajú do definovaného rozsahu pre aktuálne rozoznávanú dopravnú značku. Výsledok tejto úpravy bude obraz so zvýraznenými segmentmi. Odtieň týchto segmentov je v našom definovanom rozsahu pre tento typ značky. Nasýtenie pozadia bude znížené.



Obr. 3. Znižovanie saturácie pre farby pozadia [1].

Ďalším krokom predspracovania pre vstup optického korelátora je farebné prahovanie. To je vykonávané použitím softvérových farebných filtrov. Softvérové farebné filtre vykonávajú operáciu, ktorá odstraňuje všetky segmenty, ktoré nespadajú do zvoleného farebného rozsahu pre aktuálnu dopravnú značku. Farebnosť dopravnej značky vo video zázname závisí od podmienok okolia a od vlastností kamery. Výsledná farebná reprezentácia tej istej značky je rozdielna za jasného slnečného počasia ako napr. počas dažďa. Pre zvýšenie pravdepodobnosti správneho prahovania možno vytvoriť normalizovaný model spracovávaných farieb. Je možné redukovať vybranú farbu z 3D RGB modelu do 2D priestoru normalizáciou dvoch farieb podľa tretej [2].

IV. NÁSTROJE PRE SPRACOVANIE OBRAZU

Pre predspracovanie obrazu možno použiť knižnicu s otvoreným zdrojovým kódom s názvom Open CV [3]. Open CV je knižnica programátorských funkcií, zameraných hlavne na počítačové videnie v reálnom čase. Open CV bola pôvodne napísaná v jazyku C, no teraz má plné C++ rozhranie a celý nový vývoj je realizovaný za pomoci C++. V knižnici sa taktiež

nachádza rozhranie pre programovací jazyk Python. Open CV je vhodná na vývoj veľkého množstva typov grafických aplikácií, napr. interakcia človek - počítač, segmentácia a rozpoznávanie, sledovanie pohybu, stereo a multikamerová kalibrácia, atď.

Open CV má modulárnu štruktúru, čo znamená, že celý balíček obsahuje množstvo zdieľaných, alebo statických knižníc [3].

A. Knižnica ImgProc

ImgProc je modul Open CV pre spracovanie obrazu, ktorý zahŕňa lineárne a nelineárne obrazové filtre, geometrické obrazové transformácie, konverzie medzi farebnými priestormi, histogramy a pod. Táto knižnica obsahuje vhodné funkcie pre náš projekt. Napr. CvtColor() pre konverziu medzi farebnými priestormi a Threshold() pre aplikáciu prahovania zvoleného farebného kanála s pevnou hodnotou [4].

V. VÝPOČET VLASTNOSTÍ DOPRAVNÝCH ZNAČIEK

Dokážeme vypočítať vlastnosti značky založené na geometrickej konfigurácii kamery a značky. Pre tento účel môžeme použiť pravidlo podobnosti trojuholníkov (Obr. 4)

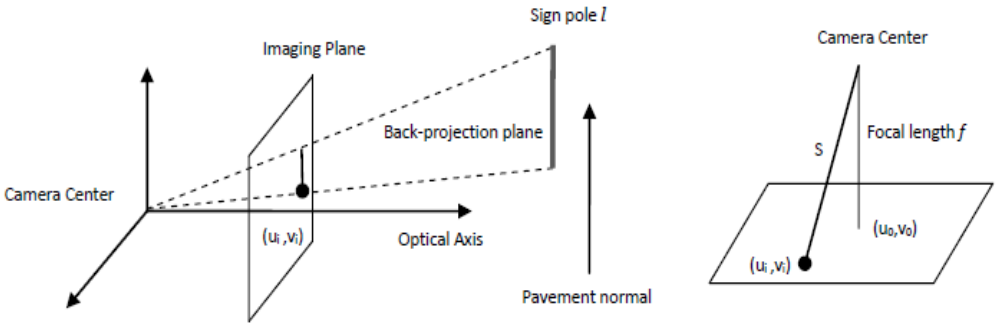
$$\frac{H}{W} = \frac{D}{S}, \tag{1}$$

kde H je výška značky, D je vzdialenosť kamery od značky. W výška značky na obraze a S vzdialenosť (v pixeloch) medzi stredom kamery a spodnou hranicou značky. S môže byť vypočítané z rozmerov obrazu, ohniskovej vzdialenosti referenčného (stredového) bodu kamery

$$S = \sqrt{f^2 + (u_1 - u_0)^2 + (v_i - v_0)^2} \tag{2}$$

Výška značky je potom vypočítaná podľa nasledujúceho vzťahu

$$H = \frac{W \cdot D}{f^2 + (u_1 - u_0)^2 + (v_i - v_0)^2} \tag{3}$$



Obr. 4. Geometrická konfigurácia dopravnej značky a kamery pre výpočet výšky dopravnej značky [2].

Touto cestou možno taktiež určiť druhý rozmer dopravnej značky. Porovnaním pomerov strán zaznamenanej značky a ideálnej značky dokážeme vypočítať horizontálny a vertikálny sklon (Obr. 4).

VI. POUŽÍVATEĽSKÉ ROZHRANIE A KOMUNIKÁCIA S DATABÁZOU

Používateľské rozhranie pre systém inventarizácie dopravných značiek môže byť vytvorený v spolupráci s Qt frameworkom [5].

Qt framework nám umožňuje písať kvalitné aplikácie a používateľské rozhrania a nasadzovať ich nezávisle od platformy bez prepisovania zdrojového kódu, čo šetrí čas a vývojové náklady.

Obsahuje množstvo knižníc pre vytváranie používateľských rozhraní, manipuláciu s

množstvom typov súborov a pre komunikáciu prostredníctvom najznámejších protokolov.

Potrebujeme taktiež úložisko pre pracovné údaje ako tvary značiek, piktogramy v značkách, zoznam nainštalovaných dopravných značiek a ich vlastnosti po vykonaní revízie. Veľmi vhodný databázový systém pre tento účel je MySQL vytvorený spoločnosťou Oracle. MySQL je voľne použiteľný a šíriteľný so všetkými typmi aplikácií. Je platformovo nezávislý a môže byť použitý ako na osobnom počítači, tak aj na samostatnom serveri. Veľkou výhodou je priama implementačná podpora v programovacích jazykoch zameraných na vývoj internetových aplikácií, čo je vhodné pre efektívne zdieľanie medzi dvomi, alebo viacerými dopravnými spoločnosťami [5].

VII. Postup rozpoznávania a inventarizácie

Na základe vyššie uvedených informácií možno popísať celý proces inventarizácie dopravných značiek:

1. Scénu vozovky zaznamenáme digitálnou kamerou.
2. Vykonáme segmentáciu a prevod do binárneho tvaru pre každú charakteristickú farbu dopravnej značky. Napríklad, značka popisujúca maximálnu povolenú rýchlosť má tvar červeného kruhu s čiernym číslom uprostred.
3. Pripravíme všetky možné ideálne tvary na porovnanie s obrazmi získanými z kroku 2.
4. Ideálne tvary a binárne vstupné obrazy spojíme do jedného binárneho obrazu pre vstup optického korelátora, postupne v závislosti od typu značky.
5. Vyhodnotíme výstup korelátora z porovnávania tvarov a určíme polohu vhodného kandidáta v obraze – korekcia polohy GPS senzorom, resp. laserovým senzorom vzdialenosti.
6. Vyhodnotíme výstup korelátora pre detekciu čísla a potvrdíme pozíciu tejto značky.
7. Z rozmerov získaných z video záznamu zistíme geometrické parametre značky.
8. Keď je značka správne rozoznaná a jej parametre sú správne, potom ju zaznamenáme do databázy. V inom prípade program zobrazí chybovú správu súvisiacu s nesprávnymi parametrami dopravnej značky.

VIII. Záver

V dôsledku zdokonaľovania technológií sú optické procesory a korelátory stále viac a viac dostupné pre komerčnú sféru. Majú veľký potenciál pre použitie v doprave a súvisiacich oblastiach, pretože dokážu vykonávať Fourierovu transformáciu v reálnom čase. Ich implementáciou je ušetrená veľká výpočtová kapacita elektronických systémov. V našom ďalšom výskume budeme vytvárať detailnejší návrh a realizáciu systému, ktorý bude schopný plne automaticky riadiť inventarizáciu dopravných značiek.

Poďakovanie

Podporujeme výskumné aktivity na Slovensku/ Projekt je spolufinancovaný zo zdrojov EÚ. Tento článok bol vypracovaný v rámci projektu "Centrum excelentnosti integrovaného výskumu a využitia progresívnych materiálov a technológií v oblasti automobilovej elektroniky", ITMS 26220120055 (50%). Táto práca bola taktiež podporovaná a realizovaná z projektu Rozvoj Centra informačných a komunikačných technológií pre znalostné systémy (ITMS kód: 26220120030) na základe podpory operačného programu Výskum a vývoj financovaného z Európskeho fondu regionálneho rozvoja (50%).

Literatúra

[1] M. Petruš, "*Optical correlator at video driver assistance system*", Diploma work, Technical University of Košice, Košice, Slovakia, 2011.
[2] Z. Hu, "*Intelligent Road Sign Inventory (IRSI) with Image Recognition and Attribute Computation from Video Log*", Kyoto, Japan, 2011.
[3] G. V. Massaguer, "*Road Sign Detection and Shape Reconstruction using Gielis Curves*", Bourgogne: Master in Computer Vision, 2011, ch. 4.
[4] *Open CV library*, published on: http://opencv.willowgarage.com/wiki/
[5] *Qt framework*, published on: http://qt.nokia.com
[6] Š. Toth, "*Rozpoznávanie dopravných značiek a ich použitie v mapových aplikáciách*", GIS, Ostrava, 2011.
[7] *Cambridge optical correlator*, published on: http://www.cambridgecorrelators.com/technology.html

Testovanie precíznych ČAP založené na porovnávaní s testovacím napätím

Martin SEKERÁK, Linus MICHAELI, Ján ŠALIGA, Jozef LIPTÁK, Marek GODLA

Katedra elektroniky a multimediálnych telekomunikácií, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

martin.sekerak@tuke.sk, linus.michaeli@tuke.sk, jan.saliga@tuke.sk
jozef.liptak@tuke.sk, marek.godla@tuke.sk

***Abstrakt* — Precízne Číslicovo Analógové prevodníky (ČAP) sú významnou súčasťou mnohých elektronických zariadení. S rastúcimi požiadavkami na množstvo spracovávaných dát ako aj rýchlosť prevodu ich výkonnosť spoločne s rozlíšením neustále rástli a to až do takej miery, že problémom sa stáva samotné testovanie. Navrhovaná metóda umožňuje meranie nelinearity ČAP za dynamických podmienok porovnávaním výstupného napätia ČAP s testovacím pomocou rýchleho komparátora a následnej registrácie vstupného kódového slova v okamihu zhody oboch napätí. Testovací signál je superpozíciou jednosmerného napätia a ditheringového signálu. Presnosť merania DC spoločne s linearitou ditheringu predurčujú metrologickú presnosť. Navrhnutý algoritmus umožňuje zo zaznamenaných kódových slov určiť integrálnu (INL) a diferenciálnu (DNL) nelinearitu. Teoretická analýza a simulácie potvrdzujúce výkonnosť navrhovanej metódy budú prezentované taktiež.**

***Kľúčové slová* — Číslicovo analógový prevodník, ČAP, dynamické testovanie, integrálna nelinearita, diferencialna nelinearita**

I. ÚVOD

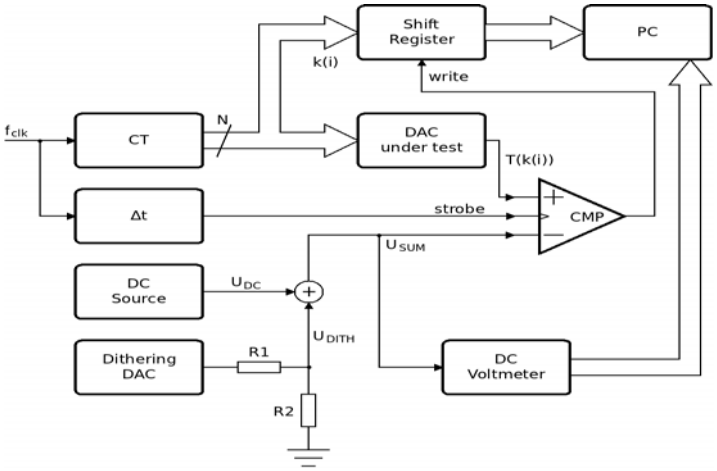
Neustále rastúce požiadavky na presnosť a rýchlosť signálových syntetizátorov, ale ja mnohých ďalších prevažne komunikačných obvodov si vyžadujú použitie rýchleho Č/A prevodníka s vysokým rozlíšením. Výroba ako aj prax si vyžadujú spoľahlivé testovacie metódy pre ich meranie realizovateľné za pomoci bežne laboratórne dostupných prístrojov. Množstvo už publikovaných testovacích metód [1] - [5] je často krát obmedzených požiadavkami na použité zariadenia a navyše by sme ich mohli prevažne charakterizovať ako statické. Dynamické testovacie metódy na druhej strane vyžadujú použitie presných spektrálnych analyzátorov alebo filtrov [6] - [9].

Táto navrhovaná metóda je založená na porovnávaní dynamického výstupného napätia Č/A prevodníka (DACuT) s testovacím signálom za pomoci rýchleho komparátora. Testovací signál je kombináciou jednosmerného referenčného DC a malého ditheringového napätia. Vstupom pre testovaný Č/A prevodník je digitálny signál pílového priebehu s možnosťou nastavovania strmosti výstupného analógového signálu. Následne digitálne spracovanie umožňuje vypočítanie vzťahu medzi načítavanými vstupnými kódovými slovami a výstupným napätím testovaného prevodníka. Vykonaním procedúry pre celý rozsah a vyhodnotením výstupných hodnôt analógového napätia testovaného Č/A prevodníka umožnuje určenie úplnej prevodovej charakteristiky a následne vypočítanie nelinearít. Presnosť testovacej metódy je najviac ovplyvnená presnosťou určenia jednosmerného DC napätia ale aj mnohými ďalšími parametrami.

Metóda poskytuje viacero výhod v porovnaní s dostupnými testovacími metódami v podobe použitia prevažne digitálneho spracovania poskytujúceho množstvo výhod a taktiež vyžaduje len presný DC voltmeter, ktoré sú relatívne cenovo dostupné. Ďalšou veľmi zaujímavou je možnosť súčasného vyhodnotenia niekoľkých kvantizačných úrovní použitím jedného referenčného napätia alebo dokonca len jedného záznamu, čo robí metódu obzvlášť zaujímavou, nakoľko efektívne redukuje čas celkového merania.

II. NAVRHOVANÁ METÓDA PRE TESTOVANIE Č/A PREVODNÍKOV

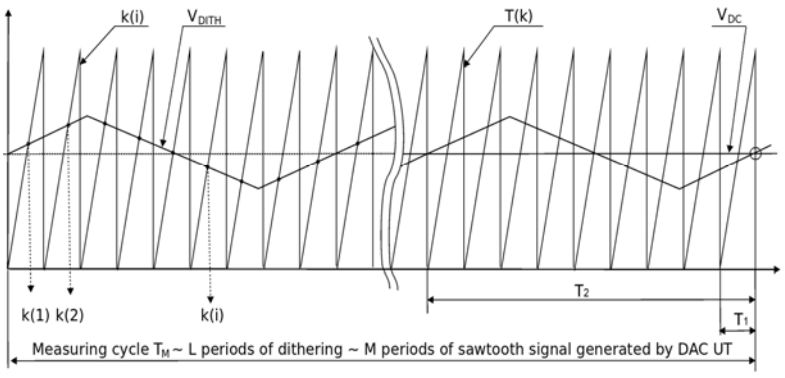
Vstupom pre testovaný Č/A prevodník je digitálny pílový priebeh a výstupom je analógové napätie ovplyvnené nelinearitami daného prevodníka, ale aj jeho ďalšími chybami. Toto výstupne napätie sa porovnáva s testovacím použitím rýchleho napäťového komparátora. Výstup komparátora následne riadi zápis vstupných kódových slov testovaného ČAP do posuvného registra v okamihu zhody oboch napätí (viď Obr. 1). Výstup komparátora je registrovaný s časovým oneskorením Δt, aby sme sa vyhli dynamickým chybám komparátora ako aj glitchom na výstupe testovaného prevodníka. Takýmto spôsobom získame pole vstupných kódových slov $k(i)$ pre M períód pílového priebehu na výstupe testovaného ČAP, ktoré sú pre zvolenú dĺžku jedného meracieho cyklu T_M následne spriemernené. Získaná hodnota spoločne s hodnotou referenčného DC napätia sú spracované navrhovaným algoritmom, ktorý vyhodnotí napäťovú odchýlku ΔU meranej kvantizačnej úrovne $T(k)$ testovaného prevodníka od referenčného napätia U_{DC} a tak umožní určenie skutočnej hodnoty výstupného napätia. Opakovaním cyklu pre celý rozsah získame úplnu prevodovú charakteristiku testovaného ČAP.



Obr. 1 Blokové zapojenie pre meranie navrhovanou metódou

Pridanie malého ditheringového napätia trojuholníkového priebehu k referenčnému DC napätiu meranému presným digitálnym voltmetrom navyše umožňuje zvýšiť presnosť merania pod hodnotu kvantizačného kroku. Počet períód ditheringu L by mal predstavovať celé číslo, aby sa spriemeroval rovnaký počet vzoriek získaných počas kladnej aj zápornej špičky. Presnosť amplitúdy ditheringu ako aj jeho nelinearity majú priamy vplyv na konečnú neistotu celej testovacej metódy. Pasívny napäťový delič $R2/(R1+R2)$ zvyšuje presnosť generovaného ditheringového napätia úmernú redukcii napätia. Metrologická presnosť navrhovanej testovacej metódy je garantovaná presnosťou použitého digitálneho voltmetra.

Navrhnutý algoritmus spracovania digitálych kódových slov potláča chyby ofsetu a vlastného oneskorenia komparátora a umožňuje kalkuláciu INL a DNL nelinearít. Postačujúcou podmienkou takéhoto merania je fakt, že časové oneskorenie komparátora je konštantné, respektíve pomaly sa meniace na celom rozsahu.



Obr. 2 Znázornenie analógového výstupného napätia $T(k)$ testovaného ČAP a testovacieho napätia daného kombináciou jednosmerného U_{DC} a ditheringového napätia U_{DITH}

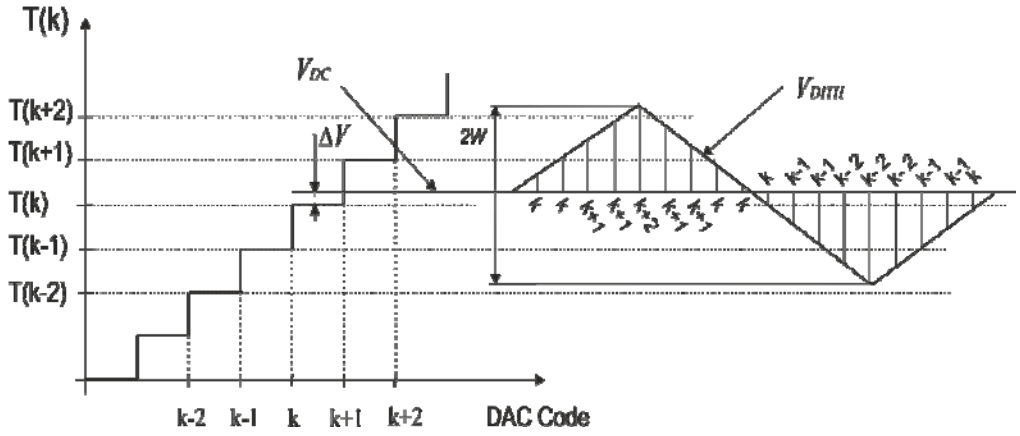
Uvažujeme, že jeden merací cyklus obsahuje *M* períód pílového priebehu generovaného testovaným ČAP a zároveň *L* períód ditheringového napätia ako je to znázornené na Obr. 2. Digitálny pílový priebeh je generovaný pripočítavaním jednotky k predchádzajúcemu kódovému slovu v každom hodinovom takte. Vzťah medzi periódou hodín a periódou jedného pílového priebehu je tak daný

$$T_1 = 2^N * T_{CLK} \tag{1}$$

Aby sme sa vyhli koherencií oboch napätí počas jedného meracieho cyklu, potom konštanty *M* a *L* by mali byt nesúdeliteľné. Túto požiadavku je možné zapísať nasledovne:

$$\frac{L}{M} = \frac{T_1}{T_2} = \frac{2^N * T_{CLK}}{T_2} \tag{2}$$

Ditheringové napätie je generované tak, že jeho amplitúda odpovedá niekonásobku LSB testovaného ČAP. Meranie jednosmerného napätia presným digitálnym voltmetrom počas jedného cyklu spriemerní takúto malú odchýlku a tak určí iba presnú hodnotu generovaného jednosmerného napätia.



Obr. 3 Ilustrácia výseku prevodovej charakteristiky testovaného ČAP spoločne so zakreslením testovacieho napätia a taktiež znázornenie zapisovania kódových slov.

Proces samotného merania vyzerá nasledovne, komparátor zdeteguje okamih, kedy výstupné analógové napätia testovaného ČAP prekročí hodnotu testovaného napätia (viď. Obr. 3). Zaregistrované digitálne kódové slovo tak splna nasledujúcu podmienku

$$T\left(k\left(i\right)\right) \geq U_{SUM} > T(k(i)-1) \tag{3}$$

Počas meracieho cyklu sa do shift registra uloží *M* hodnôt *k(i)* , *i = 1....M*, pričom ich priemerná hodnota je priamo závislá na napäťovej odchýlke *ΔU* medzi skutočným analógovým napätím a referenčným jednosmerným DC napätím meraným pomocou presného digitálneho multimetra. Na konci každého meracieho cyklu sa obsah registra spoločne s hodnotou DC napätia prenesú do počítača. Tu sa zaregistrované kódové slová *k(i)* pre testovanú kvantizačnu úroveň *k* transformujú do nového poľa $p_k(i)$

$$p_k\left(i\right) = \begin{cases} 1 & \text{if } k\left(i\right) \leq k \\ -1 & \text{if } k\left(i\right) > k \end{cases} \tag{4}$$

Pre napäťovú odchýlku *ΔU(k)* tak môžeme napísať nasledujúci vzťah:

$$\Delta U\left(k\right) = W.P_{AV} = W\left(\frac{1}{M}\sum_{1}^{M} p_k(i)\right) \tag{5}$$

kde P_{AV} predstavuje priemernú hodnotu transformovaného poľa $p_k(i)$.

Reálna hodnota výstupného analógového napätia testovaného ČAP je určená nasledujúcim vzťahom:

$$T(k) = U_{DC} - \Delta U(k) \tag{6}$$

Presné určenie hodnoty výstupného analógového napätia testovaného Č/A prevodníka následne umožňuje jednoduchý výpočet Integrálnej a Diferenciálnej nelinerity podľa nasledujúcich vzťahov:

$$INL(k) = \frac{T(k) - k * Q}{Q}$$
$$DNL(k) = \frac{\Delta U(k+1) - \Delta U(k) - Q}{Q} \tag{7}$$

Výhodou výpočtu DNL v porovnaní s výpočtom INL je fakt, že jeho hodnotu je možné určiť z jedného záznamu vytvoreného pri jednej hodnote referenčného DC napätia a zárovaň si nevyžaduje poznanie presnej hodnoty referenčného DC napätia.

Medzi ďalšie výhody patrí tiež fakt, že použitím väčšej amplitúdy ditheringového napätia ale aj samotným princípom navrhovaná metóda umožňuje vyhodnotenie viacerých kvantizačných úrovní z jedného záznamu. Takýmto spôsobom sa efektívne skráti čas potrebný pre meranie, ktorý takýmto spôsobom môže klesnúť až *2W* krát, čím by sa metóda stala zaujímavou pre nasadenie v praxi.

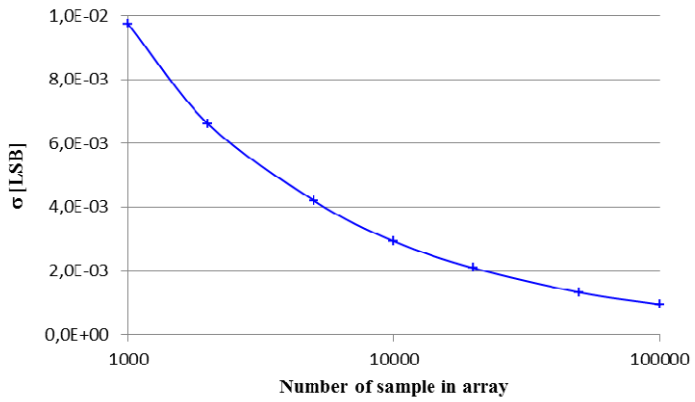
Presnosť navrhovanej metódy je ovplyvňovaná rôznymi zdrojmi chýb, ktoré zanášajú rôznu neistotu do finálneho vyhodnotenia nelinearít. Medzi takéto zdroje chýb v prípade navrhovanej metódy testovania patria:

- Chyby a neistoty referenčného DC napätia, zahŕňajúc neistoty DC generatora a tiež presného digitálneho voltmetra
- Nelinearita ditheringového napätia a chyba jeho amplitúdy *W*
- Štatistická neistota konečného záznamu $p_k(i)$
- Chyby spôsobené dynamickými vlastnosťami napäťového komparátora
- Vlastné hodnoty pracovných podmienok

III. VÝSLEDKY

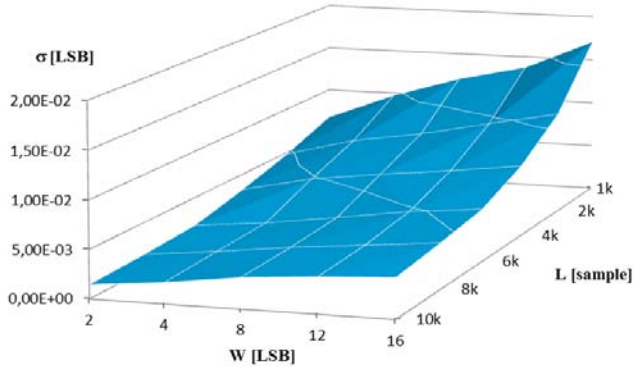
Navrhovaná metóda testovania ČAP bola simulovaná vo vývojovom prostredí LabView s cieľom potvrdiť správanie sa metódy a samotnej schopnosti presného merania. V tomto duchu boli zároveň vytvorené simulácie umožňujúce posúdiť vplyv vybraných aspektov (pracovné podmienky aj vplyvy rôznych zdrojov chýb) na celkovú presnosť merania.

Prvým takto verifikovaným parametrom je dĺžka záznamu *M*, ktorá efektívne potláča rôzne rušivé vplyvy a má tak významný vplyv na celkovú presnosť (reprezentovaná neistotou). Na nasledujúcom Obr. 4 je znázornená grafická závislosť vplyvu dĺžky záznamu *M* na celkovú neistotu vyhodnotenia integrálnej nelinearity, pričom si môžeme všimnúť jej exponenciálny pokles. S rastúcim počtom vzoriek v zázname sa čas merania neúmerne predlžuje a tak môžeme povedať, že už pri hodnote nad 20.000 vzoriek je neistota zanedbateľná a ďalšie predlžovanie záznamu neefektívne.



Obr. 4 Závislosť neistoty vyhodnotenia *INL* navrhovanou metodou v závislostí na dlžke použitého záznamu pre konštantnu amplitudu *W = 4 LSB* a *M/L=11.48*.

Ďalším významným parametrom je použitá amplitúda ditheringového napätia *W* zanašajúca neistotu do celkového vyhodnotenia integrálnej, ale aj diferencialnej nelinearity. Závislosť vplyvu amplitúdy ditheringu *W* na celkovu neistotu môžeme vidieť na Obr. 5, kde je zobrazená ako 3D grafická závislosť, kde tretím parametrom je rôzna dlžka záznamu.



Obr. 4 Závislosť neistoty vyhodnotenia *INL* navrhovanou metodou v závislostí na amplitúde použitého ditheringového napätia *W* pri rôznych dlžkach záznamu

Z grafu je vidiet, že vplyv amplitúdy ditheringu na celkovú neistotu navrhovanej metódy je ďaleko nižší ako vplyv dĺžky záznamu no je dôležité podotknúť, že vyššia amplituda ditheringu umožňuje vyhodnotenie viacerých kvantizačných úrovní súčasne.

IV. Conclusion

Článok prezentuje novú metódu testovania Č/A prevodníkov poskytujúcu množstvo výhod v porovnaní zo súčasne používanými metódami. Táto bola overená prostredníctvom simulácií, ktoré potvrdili predpokladané vlastnosti navrhovanej metódy a jej chovanie sa pre rôzne pracovné podmienky a zároveň bol overený vplyv významných zdrojov neistôt na vyslednú presnosť. Najvýznamnejšie závislosti boli prezentované a popísané. Najmä vplyv dlžky záznamu a amplitúdy použitého ditheringu si vyžadujú starostlivý výber zohľadnujúci požiadavky na presnosť a taktiež čas merania podľa potreb užívateľa.

Poďakovanie

Táto práca je súčasťou riešenia projektu VEGA, číslo 1/0555/11.

Táto práca bola vytvorená realizáciou projektu Rozvoj Centra informačných a komunikačných technológií pre znalostné systémy (kód ITMS projektu: 26220120030) na základe podpory operačného programu Výskum a vývoj financovaného z Európskeho fondu regionálneho rozvoja.

Literatúra

[1] D.L. Cari, D. Grimaldi, "Static characterization of high resolution DAC based on over sampling and low resolution ADC", *Proc. Of IEEE Instrum. And Measur. Techn. Conf., IMTC 2007*, May 1-3, 2007, Warshaw Poland

[2] L. Jin, H. Haggag, R. Geiger, and D. Chen, "Testing of precision DACs using low-resolution ADCs with wobbling," IEEE Trans. Instrum. Meas., vol. 57, no. 5, pp. 940–946, May 2008.

[3] C. W. Lin, S. F. Lin, and S. F. Luo, "A new approach for nonlinearity test of high speed DAC," in Proc. IEEE Int. Workshop Mixed-Signals, Sens., Syst., 2008, pp. 1–5.

[4] Baccigalupi, A., Liccardo, A., Grimaldi, D., Carni, D.L.: "Digital to analog converters test based on time to voltage conversion", Proc. of IEEE Instrumentation and Measurement Technology Conference, Binjiang, Hangzhou; May 2011 , Article number 5944277, pp. 6-11

[5] B. Vargha, J. Schoukens, and Y. Rolain, "Static nonlinearity testing of digital-to-analog converters," IEEE Trans. Instrum. Meas., vol. 50, no. 5, pp. 1283–1288, Oct. 2001.

[6] D.L. Carnì, D. Grimaldi, "Comparative analysis of different acquisition techniques applied to static and dynamic characterization of high resolution DAC", Proc. of XIX IMEKO World Congress "Fundamental and Applied Metrology", Lisbon, Portugal, September 6-11, 2009.

[7] J. Savoj, Ali-Azam Abbasfar, A. Amirkhany, B.W. Garlepp, M.A. Horowitz, "A new technique for characterization of Digital-to-Analog Converters in high-speed systems" Design, Automation & Test in Europe Conf.& Exhibition, 16-20 April 2007, pp.1-6.

[8] A. Baccigalupi, M. D'Arco, A. Liccardo, M. Vadursi, "Test equipment for DAC's performance assessment: design and characterization," accepted for publication on IEEE Trans. on Instrumentation and Measurement.

[9] M. D'Arco, A. Liccardo, and M. Vadursi, "Design of a test equipment for DAC's performance assessment," in Proc. IEEE I2MTC, Singapore, May 5–7, 2009, pp. 1232–1237.

Bandpass Sigma-Delta ADC Implemented on PSoC and MSI Circuits

Martin SEKERÁK, Linus MICHAELI, Ján ŠALIGA, Marek GODLA, Jozef LIPTÁK

Department of Electronics and Multimedia Communications, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Slovak Republic

martin.sekerak@tuke.sk, linus.michaeli@tuke.sk, jan.saliga@tuke.sk, marek.godla@tuke.sk, jozef.liptak@tuke.sk

***Abstract* — The article deals with the implementation of a BP ΣΔ modulator, which represents a core block of bandpass ΣΔ Analog to Digital Converters. These converters are very frequently used in software determined communication equipment as well as software radio technologies supported by manufacturers and individuals, which require the use of programmable converters of narrow-band digital modulating signal. The described BP ΣΔ ADC is an optimal tool for direct conversion of the output impedance of the two-parameter sensor into two digit flows representing its real and imaginary parts. ΣΔ Converters are suitable for the implementation in the VLSI technology, as they require the least possible amount of analog functional blocks. The implementation of ΣΔ ADC in programmable mixed signal circuits for the pre-processing and digital conversion of signals from the impedance sensors is described.**

***Keywords* — Sigma-Delta converter, Analog-Digital Converter, PCoC, pressure sensor**

I. INTRODUCTION

Recently, analog to digital converters (ADC) are the electronic blocks extremely widely used in various areas of technology, e.g. in communication systems [5], [6], computers [7], sensor technologies [1], [10] and elsewhere. There is a great number of ADC architectures with a large variety of properties and parameters [8], [9]; and another issue of study is their testing [11], [12]. Nowadays, sigma-delta ADC are more and more frequently used even in high-frequency communication systems and equipment, e.g. software radio, where a very important part of them is constituted by bandpass sigma-delta ADC [2] - [6].

The structure of the bandpass ΣΔ converter (BP ΣΔ) is based on the idea of LP ΣΔ modulator suppressing quantization noise around the converting signal frequency, with the possibility to establish balance between converting resolution and sampling period by the digital processing algorithm in the output low-pass filtering. The internal structure of both types of modulators is shown in Fig. 1.

The substitution of the low-pass filter for the resonator causes that low-pass modulator is changed to a band-pass one. The resonator is formed by the time discrete structure in the surrounding of resonation frequency f_{IN}, which is determined by the sampling frequency $f_S=4.f_{IN}$. Besides positive properties, e.g. stability and dynamics, the noise transfer function (NTF) reaches its minimum level around the sampling frequency $f_S=4.f_{IN}$ [5]. The binary multiplication with the reference signal f_{IN} shifts the spectral components at the modulator´s output around the DC component. The low-pass digital filter provides conversion of this component to the binary value. The phase shift of the reference signal is one period *Ts*, and digital value in the LP filter output representing the input signal component f_{IN} shifted by $\pi/2$.

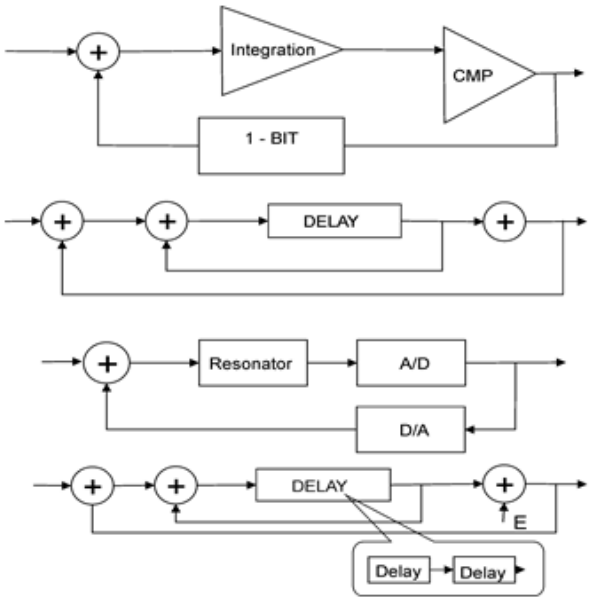


Fig. 1 LP ΣΔ modulator and its discrete model, and transformation of LP ΣΔ modulator to bandpass

The main benefit of BP ΣΔ converters is the direct conversion of the input analog signal to a vector with two components (real and imaginary). The high linearity caused by the LP filtering of the serial binary flux is a well know advantage of ΣΔ converters. The entire structure of the BP ΣΔ A/D converter is shown in Fig. 2. The resolution of each component at the digital output is adjustable by the length of low-pass FIR filtration.

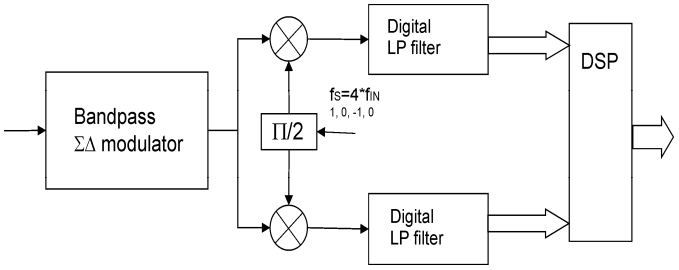


Fig. 2 Consequential AD converter for decoding I/Q components of harmonics signal.

The prospective application of BP ΣΔ converters is in the area of sensor technologies [1]. Here the signal from the two-parameter sensor is gained, where one parameter affects the real component and the other parameter affects the imaginary component of the output impedance of the passive sensor. Alternatively, one component output impedance of the sensor can represent diagnostic information about the sensor functioning.

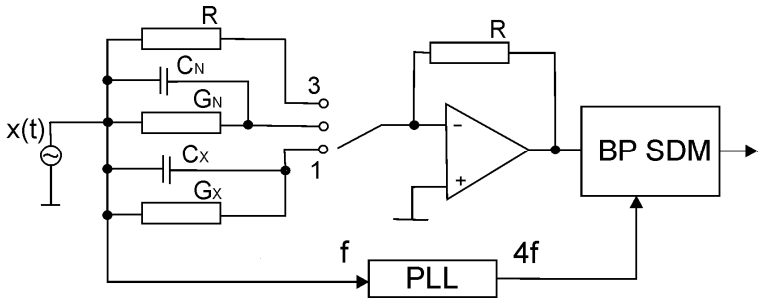


Fig. 3 Diagram of BP ΣΔ ADC as a circuit for post-processing the output signal from the capacitive sensor

The phase shift of the processing block is suppressed by the subtracting operation. The difference of the digital values from the output of the LP filter in the position "1" and "2"

normalized to the value measured at the position "3" is expressed by the formula. Here the measured changes of capacity and resistance are obtained:

$$\frac{U_1 - U_2}{U_3} = R\left(j2\pi f_s \Delta C_X + \Delta G_X\right) \tag{1}$$

The output of modulator in Z-domain, consists of the direct transfer function and superimposed noise. The power density at the output of BP ΣΔ modulator after noise shaping in the spectral domain is given by the following

$$\varepsilon_{NS}^2(f) = \varepsilon_q \frac{4}{f_S}\left(1 + \cos\frac{4\pi f}{f_S}\right) \tag{2}$$

The BP ΣΔ A-D converters can be build-in using various mixed signal technologies. They contain a minimal amount of analog elements (comparator, two integrators and summing circuits), and their offsets are compensated by the feedback structure. In addition using the simple analog preprocessing structure brings benefit in a high linearity.

II. PROGRAMMABLE SYSTEM ON CHIP (PSOC)

Circuit Programmable System on Chip (PSoC) of Cypress Microsystems allows to integrate analog and digital programmable blocks into one integrated circuit. System On Chip (SoC) consists of a microprocessor and peripheries which are controlled by the software.

The used circuit CY8C27443-24PVI is created on the basis of the SONOS technology. The microprocessor called M8C is built on the 8-bits hardware architecture, and allows effective programming. The digital part consists of FPGA blocks organized in rows. Analog blocks are used for the development of analog elements, e.g. comparators, summing amplifiers as well as AD and DA Converters. The configuration and emulation of the circuit are performed in the PSoC Designer graphical development environment.

The analog part contains three types of the analog blocks: Continuous Time (CT) which provides continuous time analog functions. Two other blocks represented by the Switching Capacitor blocks (SC) of C and D type provide programmable signal amplification or integration using the switching capacitor technology, as shown in Fig. 4. Blocks are organized in columns, where some of them have the possibility to connect their output into the inputs of the other ones. The state of the interconnection switches is stored in the mapped memory. This Flash memory contains information enabling or disabling internal switches, which are controlled by two internal phase shifted clock signals; and selection/activation of the input multiplexers.

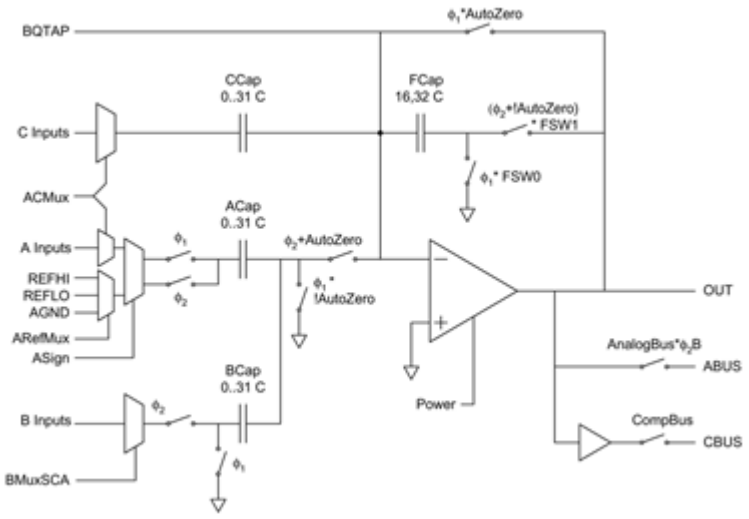


Fig. 4 Analog switched capacitor block type C

III. DRAFT BP SD MODULATOR CIRCUIT FOR PSOC

The resonator of ΣΔ modulator was designed by supplementing a double-delay circuit of the negative feedback circuit and summing circuit in the analog part of PSoC. The input analog signal was connected to the input amplifier together with the signal from the comparator´s output through a resistive summing circuit. The output signal then passes the double-delay resonator. The double-delay circuit was implemented by means of two SC blocks, each of them represents an amplifier with the amplification equal to one, and creating delays of one f_S clock. The transmission characteristics of the double-delay circuits with the used unity amplification are shown in Fig. 5. The effect of the SC block saturation deteriorates characteristics of higher input amplitudes.

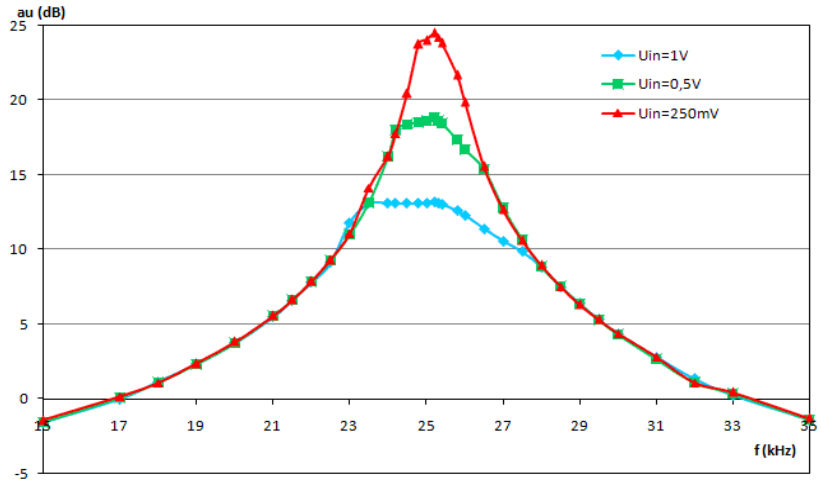


Fig. 5 Transfer characteristics of a resonator

The comparator is connected to the resonator´s output. The comparator output was a digital bus connected to the buffer located in the digital circuit. The configuring BP ΣΔ modulator in analog parts of the PSoC circuit created in the graphical development environment is shown in Fig. 6.

The input signal range was restricted by supply voltage of 0 - 5V. In order to remove unilateral restrictions, the DC bias of 2.5V was added to the input signal.

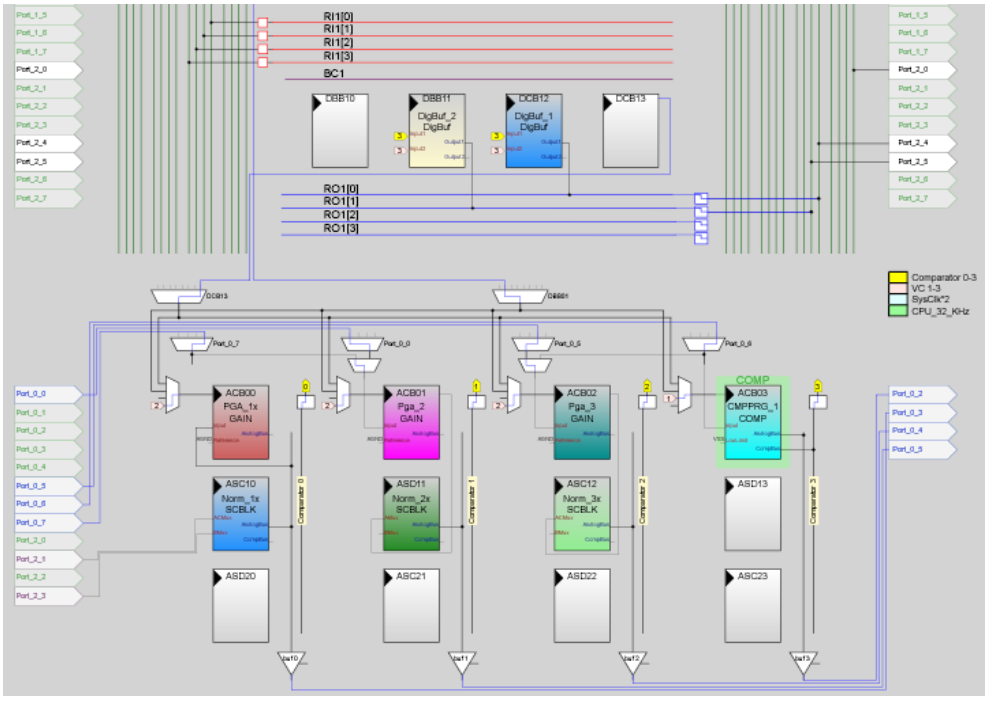


Fig. 6 Diagram BP ΣΔ modulator in the development environment PSoC Designer

IV. IMPLEMENTATION OF BP ΣΔ CONVERTER USING DISCRETE COMPONENTS

The main drawbacks of the PSoC are the limited amount of analog blocks and the fact that these blocks have a symmetrical structure resulting in a small utilizable input signal range; moreover they require DC offset setting. The BP ΣΔ modulator was also implemented using MSI components.

The delay-block z^{-1} was formed by a pair of sampling amplifiers (SA). The non-overlapping clock with 50% duty cycle on each SA in one block ensures time delay z^{-1}. Comparators and amplifiers were implemented on broadband integrated circuits. The resistive summing circuit maintains an accurate summing operation without a drift. The final circuit structure of BP ΣΔ modulator implemented on MSI circuits is shown in Fig. 7.

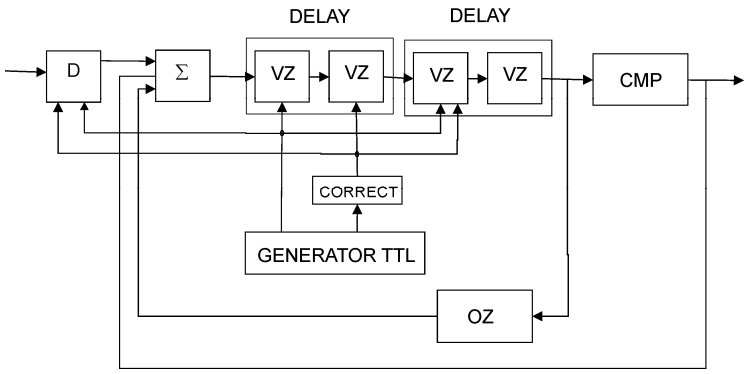


Fig. 7 Structure bandpass ΣΔ modulator formed by MSI circuits

Acquisition of the binary output from the BP ΣΔ modulator was performed by DAQ NI USB-6521 for both circuit implementations. Initially the acquired binary flux from the BP ΣΔ modulator was multiplied by two sequences of the binary sampling signal mutually shifted by one tact. This sampling shift corresponds to the phase shift $\pi/2$ of the input signal f_{IN}. The successive signal LP filtering was performed by a hosting PC, where the program was created in a LabView development environment by NI.

V. EXPERIMENTAL RESULTS

In both implementations of the created BP ΣΔ modulator, the measurement of structures was performed by the oscilloscope, and simultaneously by the DAQ card and binary data registration program from the modulator created in LabView. The measured course of the BP ΣΔ modulator output value for a single value of the input voltage U_{IN}=0.5V on the oscilloscope record is shown in Fig. 8, while the resulting binary sequence is shown in Fig. 9.

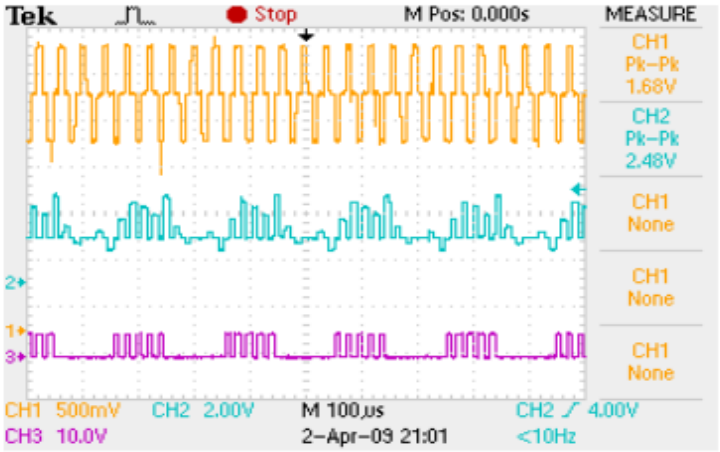


Fig. 8 The measured course on the BP ΔΣ modulator output from the oscilloscope for the input signal amplitude U_{IN}=0. 5V

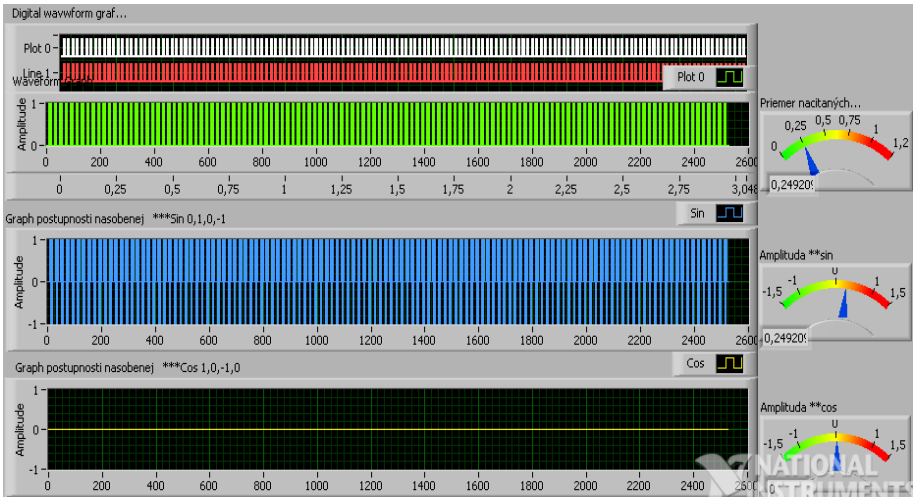


Fig. 9 Course and numerical evaluation of the binary data from the BP ΣΔ modulator output for amplitude of the input signal U_{IN} =0.5V obtained by means of DAQ card and the binary data registration program.

The multiplication by the binary sampling frequency moved the noise shaped signal from BP ΣΔ modulator to DC at the frequency axis. The resulting signal was filtered by a digital low-pass filter. The transfer characteristics for SoC as a function of the feedback resistance can be seen in Fig. 10. The linearity of the transfer function between the input signal amplitude and the digital output was limited. The slope of the transfer characteristics and the input signal range can be controlled by the resistance in the BP ΣΔ modulator feedback.

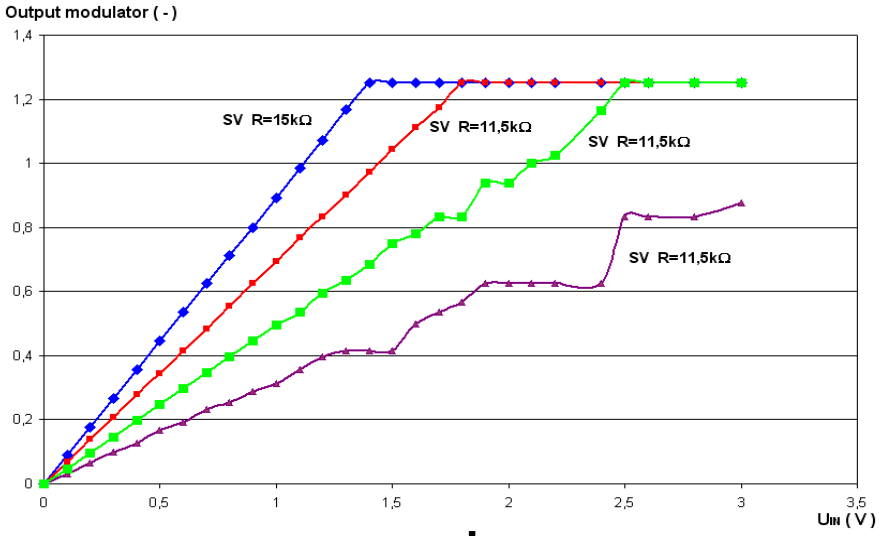


Fig. 10 The resulting transfer characteristics of BP ΣΔ modulator implemented on the PSoC circuit

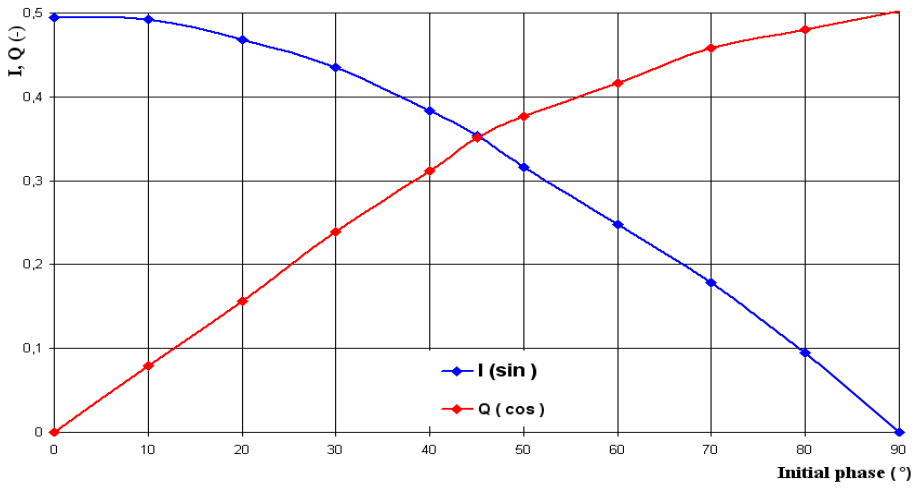


Fig. 11 Dependency of I and Q components on the initial phase of the analog input signal

One of the major characteristics of this converter is its sensitivity to the input signal phase, as reflected by changing values of I and Q components obtained by the processing of binary information from the modulator. Fig. 11 shows the dependence of I and Q components on the phase of the input analog signal, which was converted.

VI. CONCLUSION

Compared with other types of converters, ΣΔ modulators offer many advantages including simple feasibility since they are based on the digital processing using minimal amount of analog blocks. SoC belongs to the group of mixed signal circuits, where both digital and analog part is programmable according to user´s needs. Especially digital multiplication and low-pass filtering can be performed on the same chip as analog preprocessing operations. The main shortcoming of this solution is a small range of the transfer characteristics linearity and restricted possibility to perform analog operations in the differential mode. The additional offset causes thermal drifts in the analog preprocessing blocks.

The implementation of ΣΔ modulator on MSI components reduces this drawback to the detriment of the consistent block structure with both analog and digital integrated circuits. The strength of this implementation is the possibility to modify any part independently. For users with available design tools ASCI circuits represent the most prospective solution.

ACKNOWLEDGMENT

The work is a part of project supported by the Science Grant Agency of Slovak republic (No. 1/0555/11).

This work was supported by Agency of the Ministry of Education of the Slovak Republic for the Structural Funds of the EU under the project Development of Centre of Information and Communication Technologies for Knowledge Systems (project number: 26220120030).

REFERENCES

[1] Haze J., Vrba R., Fujcik L., Forejtek J., Zavoral P., Pavlik M., Michaeli L.: "The Novel Band-Pass Sigma-Delta Modulator Utilised for Capacitive Pressure Sensing" ICONS '7 Second International Conference, 22 – 28 April 2007

[2] Krukowski A., Kale I., Morling R. : "Applications of polyphase filters for bandpass Sigma-Delta analog-to-digital conversion", International Workshop on ADC Modeling, May 1996

[3] C. I. Lao, H. L. Leong, K. F. Au, K. H. M and S. P. U, "A 10.7-MHz bandpass ΣΔ modulator using double-delay single-opamp SC Resonator with double-sampling", IEEE Circuit and System, May 2003, pp 1061-1064.

[4] Brigati S., Francesconi F., Malcovati P., Tonietto A., Maloberti F.: "Modeling Sigma-Delta Modulators Non-Idealities in SIMULINK", Proceedings of IEEE International Symposium on Circuit and Systems, Orlando USA, 1999

[5] Salo T., "An 80-MHz bandpass ΔΣ modulator for a 100-MHz IF receiver", Solid-State Circuits, IEEE Journal of, vol. 37, pp. 798-808, 2002.

[6] H. K. C. Issac, "A 70 MHz CMOS band-pass sigma-delta analog-to-digital converter for wireless receivers", M.S. thesis, Hong Kong University of Science and Technology, August 1999

[7] ŠALIGA, Ján - MICHAELI, Linus: Software for metrological characterisation of PC sound cards. In: Computer standards and interfaces. vol. 25, no. 1 (2003), p. 45-55

[8] MICHAELI, Linus - ŠALIGA, Ján - KOLLÁR, Martin: Parameters of band pass sigma delta ADC and the comparison with the standard ones. In: Measurement. vol. 40, no. 5 (2007), p. 473-478.

[9] ŠALIGA, Ján: Testovanie parametrov analógovo-číslicových prevodníkov dynamickými metódami. In: Slaboproudý obzor. roč. 59, č. 1 (2002), s. p6-p11.

[10] [10] KOLLÁR, Martin - ŠALIGA, Ján: The principle of new sigma delta modulation technique based upon the use of a flip-flop. In: 13th International Symposium on Measurements for Research and Industry Applications: Athens - Greece, 2004. ISBN 960-254-644-1.

[11] CHOVANEC, Marián - MICHAELI, Linus - ŠALIGA, Ján: Noise shaping structures of band pass sigma delta ADC and their impact on metrological parameters. In: 16th IMEKO TC4 International Symposium, Florence, Italy, 2008. p. 977-982.

[12] A.C. Serra, F. da Silva, P.M. Ramos, L. Michaeli and J. Saliga, Combined spectral and histogram analysis for fast ADC testing, IEEE Transactions on Instrumentation and Measurement 54 (2005), pp. 1613–1617.

Analýza morfémových modelov slovenského jazyka v systémoch rozpoznávania plynulej reči

Ján STAŠ, Daniel ZLACKÝ, Jozef JUHÁR

Katedra elektroniky a multimediálnych telekomunikácií,
Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach,
Park Komenského 13, 042 00 Košice, Slovenská republika

jan.stas@tuke.sk, daniel.zlacky@gmail.com, jozef.juhar@tuke.sk

Abstrakt **– Flektívnosť slovenského jazyka prináša so sebou veľké množstvo ohybných tvarov slov, s ktorými narastá nielen veľkosť slovníka, ale aj percento mimoslovníkových slov. Morfémové jazykové modely sa snažia tento problém riešiť dekompozíciou slov do menších subslovných jednotiek reči, nazývaných morfy, a tým riešia aj všeobecný problém nedostatočnosti trénovacích dát. V tomto príspevku budú opísané viaceré prístupy na automatickú segmentáciu slov do morfém obsiahnutých v korpuse textových dát s využitím rôznych pravidlovo-založených a dátami-riadených princípov. Segmentované textové dáta boli neskôr použité pri tvorbe modelov slovenského jazyka pre systémy rozpoznávania plynulej reči v úlohe automatického prepisu spravodajských relácií. Priebežné výsledky ukazujú výrazné zníženie počtu mimoslovníkových slov a hodnôt perplexity morfémových modelov s miernym zhoršením presnosti systému na rozpoznávanie reči ako celku.**

Kľúčové slová **– automatická segmentácia slov, modelovanie jazyka, morfológia slovenčiny, rozpoznávanie plynulej reči**

I. Úvod

Bohatá morfológia a relatívne voľná stavba vetných štruktúr v prípade vysoko-flektívnych a aglutinačných jazykov spôsobuje rad nevýhod použitia štandardných n-gramových modelov pri štatistickom modelovaní jazyka v systémoch automatického rozpoznávania plynulej reči s veľkým slovníkom (v angl. „*large vocabulary continuous speech recognition*“, skr. LVCSR), ktoré vyplývajú priamo z ich podstaty. Veľký počet slovných tvarov má za následok *vzrast počtu* n*-gramov* v jazykovom modeli, a tým aj veľkosti výsledného modelu, čo má výrazný vplyv na časové a pamäťové nároky aplikácie v procese dekódovania vyslovenej postupnosti slov systémom LVCSR. Ďalšou nevýhodou celoslovných modelov jazyka je aj *nedostatočnosť trénovacích dát* (v angl. „*data sparsity*“), čo spôsobuje, že odhad podmienenej pravdepodobnosti málopočetných javov v n-gramovom modeli je často menej vierohodný. Z toho dôvodu boli navrhnuté viaceré spôsoby eliminácie nevýhod n-gramových modelov, s ohľadom na morfológiu daného jazyka. Jednou z možností je uprieť pozornosť na menšie plnovýznamové jednotky reči, akými sú morfy a ich štatistickým modelovaním reprezentovať jazyk v systéme LVCSR.

Morfémové modely (v angl. „*morph-based*“ alebo „*subword-based models*“) riešia problém nedostatočnosti trénovacích dát dekompozíciou slov do menších subslovných jednotiek reči. Takýto spôsob reprezentácie slov výrazne redukuje veľkosť slovníka a počet mimoslovníkových tvarov slov (v angl. „*out-of-vocabulary*“, skr. OOV), keďže počet všetkých možných morf je vždy menší než počet všetkých možných slov v danom jazyku a priemerný výskyt týchto subslovných jednotiek v trénovacom textovom korpuse je vždy väčší než priemerný výskyt slov v tom istom korpuse. Vo všeobecnosti, morfy môžu byť tvorené *slabikami*, *predponami*, *slovnými kmeňmi*, *koreňmi*, *príponami*, alebo *koncovkami slov*, príp. inými plnovýznamovými časťami slova, označovanými tiež *afixy*. Výber vhodných subslovných jednotiek pri štatistickom modelovaní jazyka potom záleží od viacerých parametrov, napr. od priemernej dĺžky slov v jazyku, hranicami ohybu a spôsobom tvorenia slov, od ich výslovnosti, či vnútroslovného a medzislovného pokrytia.

Oblasť modelovania jazyka pomocou subslovných jednotiek je v súčasnosti zameraná na využitie troch základných princípov reprezentácie morf v jazykovom modeli, a to na *modely založené na slabikách*, ktoré boli úspešne aplikované napr. pri modelovaní poľského jazyka [1], na *modely*

založené na gramatických a štatistických morfách, ktoré sú uplatniteľné najmä v aglutinačných typoch jazyka, akými sú napr. fínčina alebo estónčina [2] a na *modely v tvare slovného kmeňa a koncovky*, ktoré dekomponujú slová len na dve časti a sú pomerne často využívané pri modelovaní flektívnych jazykov, akými sú napr. čeština [3], slovinčina [4], alebo ruština [5].

Ako to už bolo spomenuté v predchádzajúcom odseku, podrobným štúdiom flektívnych jazykov sa dospelo k poznatku, že ohyb v danom type jazyka je v prevažnej väčšine sústredený na hranici slovného kmeňa a koncovky. Takýmto spôsobom sa výrazne zvýši aj prediktívna schopnosť jazykového modelu oproti predchádzajúcim prístupom, pri ktorých dĺžka histórie morf v rámci jedného slova všeobecne nie je pre n-gramový model nikdy známa a ten môže ich reťazením vygenerovať množstvo gramaticky nesprávnych tvarov. Z toho dôvodu sme oblasť modelovania slovenského jazyka v tomto článku upreli na analýzu a využite morfémových modelov tvorených slovnými kmeňmi a koncovkami slov. Tento spôsob je určitým kompromisom medzi veľmi malými segmentami slova, akými sú slabiky a medzi celými slovami. Krátke segmenty tiež vyžadujú n-gramy vyššieho rádu, s ktorými vzrastá aj veľkosť a perplexita výsledného modelu.

V nasledujúcich častiach tohto príspevku budú v krátkosti zhrnuté jednotlivé princípy a postupy používané pri automatickej segmentácii slov do subslovných jednotiek s využitím pravidlovo-založených a dátami-riadených štatistických metód a princípy, ktoré boli doposiaľ vytvorené na segmentáciu slov v slovenčine. V tejto časti predstavíme tiež vlastný algoritmus segmentovania slov na kmene a koncovky, ktorý sme navrhli v rámci tejto práce. Ďalšia kapitola opisuje jednotné nastavenie systému LVCSR použité pri tvorbe experimentov, o výsledkoch ktorých podáva prehľad piata kapitola tohto článku. V záverečnej časti príspevku budú spomenuté hlavné prínosy s ďalším smerovaním v oblasti tvorby morfémových modelov slovenského jazyka.

II. Metódy automatickej segmentácie slov

Ako už bolo spomenuté v úvode článku, metódy automatickej segmentácie slov do subslovných jednotiek sa vo všeobecnosti delia do dvoch základných skupín, a to na *pravidlovo-založené prístupy*, ktoré dekomponujú slová pomocou vopred zadaných pravidiel segmentácie navrhnutých pre konkrétny jazyk a na *dátami-riadené prístupy*, ktoré vychádzajú buď zo štatistických metód alebo z určitých zákonitostí daného jazyka, príp. na metódy kombinujúce oba tieto prístupy.

Jedným z najjednoduchších dátami-riadených prístupov je *sémanticky-orientovaná segmentácia slov* navrhnutá pre flektívne jazyky, ktorá vychádza z princípu *skrytej sémantickej analýzy* (v angl. „*latent semantic analysis*“), u ktorej hranice segmentácie sú dané hranicami vetvenia sa slov v stromovej štruktúre pre skupinu slov s podobnými vlastnosťami (kandidátnymi afixami) [6].

Medzi najznámejšie princípy segmentácie slov bez poznania jazyka (v angl. „*knowledge-free morphology segmentation*“), ktorá je základom aj viacerých iných dátami-riadených metód, patrí *metóda založená na rôznorodosti nasledujúcich znakov* (v angl. „*letter successor variety*“, skr. LSV), navrhnutá dr. Harrisom. Tento typ segmentácie slov je založený na výpočte frekvenčného výskytu rôznych znakov počítaných za (alebo pred) skupinou znakov (osobitne) v danom slove, pričom morfematická hranica sa nachádza v miestach za (alebo pred) maximálnou (alebo minimálnou) hodnotou početnosti rôznych znakov na danej pozícii [7].

Bordag následne rozšíril segmentáciu založenú na LSV aj o kombináciu ďalších hodnotiacich mier zahrňujúcich kontext jednotlivých znakov, a to o *frekvenciu výskytu znakov* na danej pozícii nachádzajúcich sa v slovníku, o *faktor prekrývania sa znakov*[1] v prípade prekrývajúcich sa afixov, o *multiznaky*[2], ktoré v jazyku odrážajú len jedno písmeno abecedy a o *váhu bigramu*, t. j. o pravdepodobnosť výskytu dvojice nasledujúcich, resp. predchádzajúcich znakov, osobitne [8]. Možno poznamenať, že aj keď táto metóda zvažuje viacero hodnotiacich mier na určenie správnej segmentácie slov, oproti segmentácii založenej na LSV sa konečným výsledkom príliš nelíši.

Podobný princíp dátami-riadenej segmentácie bez dohľadu bol uplatnený aj pri *metóde minimálnej opisnej dĺžky* (v angl. „*minimum description length*“, skr. MDL), prvýkrát opísanej v [9]. Tento pomerne zložitý algoritmus zvažuje model, ktorý by čo najjednoduchším a najkratším spôsobom vedel opísať skúmaný jazyk, morfologické regulárnosti v danom jazyku, ako aj samotný model jazyka, pomocou pravdepodobnosti rozdelenia jednotlivých morf v skupine slov s podobnými vlastnosťami (kandidátnymi koreňmi). Algoritmus MDL bol neskôr úspešne aplikovaný v systéme zvanom Morfessor pri modelovaní fínskeho jazyka pomocou štatistických morf [10], neskôr uplatnený aj v prípade modelovania slovinčiny [4], aglutinačnej estónčiny, flektívnej turečtiny a konverzačnej arabčiny pomocou morfémových a morfologicky-založených modelov [11].

Z ďalších prístupov možno spomenúť segmentáciu s využitím morfologického analyzátora aplikáciou *dvojúrovňovej morfologickej analýzy*[3] v prípade modelovania českého jazyka pomocou

[1] V slovenčine sú to napr. predpony „*na*“ a „*nad*“, „*po*“ a „*pod*“ a pod.
[2] V slovenčine sú to spoluhlásky „*ch*, *dz*, *dž*“ a dvojhlásky „*ia*, *ie*, *iu*, *ii*“.
[3] Priradenie slovnej lemy a hľadanie koncovky slova na základe morfologickej značky.

kmeňov a koncoviek slov [3] a štatistický prístup na určovanie slabičných hraníc slov nástrojom Sylseg pomocou *skrytých Markovových modelov* (v angl. „*hidden Markov models*“, skr. HMM), ktorý bol trénovaný na množine textových dát segmentovaných na základe jednoduchých pravidiel na slabičnú segmentáciu slov v slovenčine [12].

III. Segmentácia slov v slovenčine

Vybrané spôsoby automatickej segmentácie slov spomenuté v predchádzajúcej časti boli aplikované na slovenský jazyk s cieľom výberu tej najvhodnejšej segmentácie pre potreby modelovania slovenského jazyka pomocou subslových jednotiek.

Z výsledkov zhromaždených v [13] a [14] možno konštatovať, že segmentácia založená na výpočte LSV podľa Harrisa a jej modifikovaná verzia navrhnutá Bordagom len v malom počte prípadov dosiahla správne delenie. V množine segmentovaných slov bola správne oddelená len ich predpona, naopak veľká časť slov obsiahnutých v slovníku ostala nesegmentovaná, čo obmedzuje použiteľnosť týchto metód pri tvorbe morfémových modelov slovenského jazyka.

Ako už bolo spomenuté v úvodnej časti tohto článku, v prípade flektívnych jazykov sa ako najvhodnejší spôsob javí deliť slová na dve časti, kmeň a koncovku slova. Takýto spôsob delenia slov sme prvýkrát použili aj pri tvorbe modelov založených na triedach slov odvodených od ich koncoviek. V tomto prípade sme aplikovali jednoduchý štatistický princíp založený na výpočte frekvenčnej distribúcie dvoj- až päť-znakových koncoviek a segmentácii tej najdlhšej koncovky s najpravdepodobnejším výskytom v slovníku [15].

Keďže daný spôsob segmentácie slov zvažoval delenie len na štatisticky-zvolených morfematických hraniciach bez ohľadu na správnu fonetickú výslovnosť jednotlivých morf po segmentácii, bolo preto potrebné vytvoriť nové pravidlá, ktoré by brali v úvahu delenie slov nielen na morfematických, ale aj na fonetických (na tzv. morfo-fonologických) hraniciach vnútri jednotlivých slov. Tieto pravidlá možno zhrnúť do nasledujúcich bodov [14]:

- pravidlá na segmentáciu slov na kmene a koncovky vychádzajú z pravidiel slabičnej segmentácie v slovenčine [12], kde každá koncovka je reprezentovaná poslednou slabikou slova;
- segmentované sú len slová dĺžky viac než dve slabiky, resp. sedem písmen, zvyšné slová nepodliehajú pravidlám segmentácie;
- dĺžka koncovky je v rozsahu dvoch až štyroch písmen, resp. dĺžky jednej slabiky;
- segmentácia slov je obmedzená slovníkom, ktorý obsahuje približne 1 milión gramaticky správnych tvarov slov z databázy Slovenského národného korpusu [16].

Iný typ segmentácie slov na kmene a koncovky sme dosiahli úpravou segmentácie pomocou nástroja Morfessor [10], tak aby spĺňal jednotlivé body navrhnutej segmentácie slov uvedené v predchádzajúcom odseku. To znamená, že segmentácii nepodliehali slová, od ktorých by bola nástrojom Morfessor oddelená potenciálna predpona, príp. jednoznaková koncovka, tiež slová, ktorých dlžka bola menej než 7 znakov. V prípade, že koncovka pozostávala z viacerých segmentovaných častí, boli tieto zlúčené do jednej koncovky.

Čo sa týka slabičnej segmentácie pomocou nástroja Sylseg, daný spôsob sme neskôr rozšírili o novú trénovaciu množinu, s ktorou vzrástli pamäťové nároky pôvodného algoritmu. Z toho dôvodu bolo potrebné upraviť existujúci algoritmus využitím metódy binomického stromu, čím sa neporovnateľne skrátil čas potrebný na lokalizáciu morf v natrénovanom modeli. Upravený nástroj dostal názov Morseg a výsledky segmentácie slov pomocou tohto nástroja sú publikované v [13]. Napriek tomu sme však pri tvorbe morfémových modelov založených na slabikách použili pôvodný nástroj, ktorého aplikácia bola obmedzená dĺžkou spracovaných slov.

IV. Experimentálne nastavenie systému LVCSR

Pri experimentoch sme sa uvažovali bigramové, trigramové a štvorgramové morfémové modely, vytvorené pomocou programovej sady SRILM [17], ktoré boli obmedzené slovníkmi veľkosti $25k$ až $125k$ najviac frekventovaných slov a morf, s krokom $25k$. Štatistiku počtu jednotlivých druhov morf a slov obsiahnutých v slovníkoch pre rôzne druhy segmentácie je možné vidieť v Tabuľke 1. Pri vyhladzovaní jazykových modelov sme použili štandardný Wittenov-Bellov algoritmus v ústupovej schéme. Všetky modely boli trénované na textovej databáze novinových článkov v rozsahu približne 180 miliónov tokenov, získaných použitím nami vytvoreného systému na dolovanie, organizáciu a predspracovanie textových dát zo slovenských webstránok v rokoch 2007 až 2011, nazývaného webAgent [18].

Experimenty boli vykonávané z trifónovým kontextovo-závislým akustickým modelom vytvoreným na báze HMM, pričom každý stav HMM bol modelovaný zmesou 32 Gaussových funkcií pravdepodobnosti. Akustický model bol vytvorený z príznakových vektorov, ktoré obsahovali 39 mel-frekvenčných kepstrálnych koeficientov (v angl. „*mel-frequency cepstral coefficients*“). Natrénovaný bol na anotovaných rečových nahrávkach profesionálneho diktátu s prvkami spontánnej

Tabuľka 1
Štatistika počtu morfém a slov v jednotlivých slovníkoch a počtu n-gramov v morfémových modeloch

	počet							
ID	**slabík**	**kmeňov**	**koncoviek**	**slov**	**1-gramov**	**2-gramov**	**3-gramov**	**4-gramov**
referencia								
25k	-	-	-	25 025	25 025	10 586 981	12 049 195	12 553 487
50k	-	-	-	50 107	50 107	14 909 796	14 413 353	14 895 535
75k	-	-	-	75 271	75 271	17 335 302	15 488 378	15 957 574
100k	-	-	-	100 558	100 558	18 904 588	16 087 874	16 546 323
125k	-	-	-	125 930	125 930	20 001 665	16 458 378	16 907 362
navrhnutá segmentácia v tvare „kmeň-koncovka“								
25k	-	13 506	1 449	10 053	25 007	9 058 238	15 974 486	21 328 014
50k	-	28 395	2 062	19 777	50 234	11 422 752	17 746 052	23 273 406
75k	-	43 759	2 523	29 224	75 506	12 407 913	18 303 532	23 846 229
100k	-	59 344	2 883	37 905	100 132	12 881 029	18 520 063	24 058 365
116k	-	70 315	3 013	43 410	116 738	13 070 275	18 591 928	24 126 834
navrhnutá segmentácia v tvare „kmeň-koncovka“ - upravený slovník								
25k	-	12 556	3 013	9 432	25 011	8 868 379	15 788 594	21 112 551
50k	-	27 714	3 013	19 343	50 070	11 375 736	17 715 301	23 240 031
75k	-	43 263	3 013	28 920	75 196	12 392 334	18 295 814	23 838 079
100k	-	59 344	3 013	37 905	100 262	12 882 122	18 520 599	24 058 812
116k	-	70 315	3 013	43 410	116 738	13 070 275	18 591 928	24 126 834
Morfessor - segmentácia v tvare „kmeň-koncovka“								
25k	-	5 255	1 295	18 459	25 006	10 747 833	14 471 208	16 772 987
50k	-	10 191	2 179	37 795	50 162	14 456 810	16 704 651	19 131 599
75k	-	14 802	3 097	57 137	75 033	16 287 252	17 574 644	20 030 011
100k	-	18 971	4 032	77 352	100 352	17 406 630	18 026 104	20 485 360
125k	-	22 857	4 944	99 171	126 969	18 179 326	18 300 023	20 756 329
Morfessor rozšírený o navrhnutú segmentáciu v tvare „kmeň-koncovka“								
25k	-	12 689	2 316	10 008	25 011	9 499 969	16 278 558	21 529 386
50k	-	25 538	3 844	20 759	50 139	12 033 342	18 111 808	23 553 770
75k	-	36 875	5 350	32 810	75 033	13 155 305	18 724 889	24 191 084
100k	-	47 217	6 757	47 649	101 621	13 830 857	19 022 524	24 487 627
125k	-	54 921	7 848	62 816	125 583	14 224 152	19 168 684	24 630 897
Sylseg - slabičná segmentácia								
25k	4 685	-	-	20 405	25 090	6 726 410	14 370 827	22 684 457
50k	5 341	-	-	45 418	50 759	7 533 955	14 667 211	22 982 971
62k	5 452	-	-	56 936	62 388	7 634 265	14 688 149	23 003 215

reči v rozsahu približne 60 hodín z databázy spravodajských relácií v slovenčine, zaznamenanej v rokoch 2007 až 2009. Databáza je charakteristická vyváženosťou pohlavia rečníkov, obsahuje čítaný a spontánny prejav a v malej miere aj segmenty telefónnej reči. Nepozorované trifóny boli modelované použitím efektívneho algoritmu mapovania trifón opísaného v [19].

Ako dekodér bol použitý vysoko-výkonný rozpoznávací systém Julius, založený na dvojpásmovej prehľadávacej stratégii, ktorý v prvom prechode rozpoznávania reči využíva vždy ľavo-pravý bigramový jazykový model a v druhom prechode sú tieto hypotézy po rozpoznaní dotvorené pravo-ľavým (reverzným) n-gramovým modelom najvyššieho rádu [20].

Testovaciu množinu rečových nahrávok tvorili náhodne vybraté segmenty reči zo spomenutej databázy spravodajských relácií v rozsahu približne 240 minút. Vybrané segmenty neboli zahrnuté do trénovania akustického modelu a obsahujú celkovo 40 656 slov v 4 343 vetách.

Na hodnotenie kvality n-gramových modelov sme použili dve štandardné miery, intrinzické hodnotenie pomocou zložitosti (v angl. „*perplexity*“, skr. PPL) modelu a extrinzické pomocou miery výskytu chýb (v angl. „*word error rate*“, skr. WER), vyčísľovaných na množine testovacích dát. Perplexita modelu je definovaná ako reciproká hodnota váženého (geometrického) priemeru pravdepodobnosti pridelenej jazykovým modelom pre každé slovo v testovacích dátach. Naopak WER ohodnocuje systém LVCSR ako celok a udáva sa ako pomer hypotéz textu po rozpoznaní, ktorý v sebe zahŕňa chyby vložením, zámenou a chyby zmazaním, k referenčnému textu.

V. Experimentálne výsledky

V rámci analýzy morfémových modelov sme sa zamerali na hodnotenie ich kvality pomocou percentuálnej hodnoty počtu mimoslovníkových tvarov, perplexity modelu a chybovosti systému LVCSR, vyhodnocovaných na množine testovacích dát, s cieľom zistiť vplyv jednotlivých druhov segmentácie slov na celkovú úspešnosť systému LVCSR ako celku, v úlohe prepisu spravodajských relácii v slovenčine. Navrhnutý spôsob segmentácie slov na kmene a koncovky sme porovnávali s modifikovanou verziou segmentácie pomocou nástroja Morfessor, taktiež so segmentáciou spájajúcou oba tieto prístupy, ďalej so slabičnou segmentáciou vytvorenou nástrojom Sylseg, a to v úlohe modelovania slovenského jazyka pomocou bi-, tri- a štvorgramových morfémových

Tabuľka 2
Hodnotenie kvality n-gramových morfémových modelov pre jednotlivé druhy segmentácie slov

ID	veľkosť slovníka	OOV [%]	2-gramy PPL	2-gramy WER [%]	3-gramy PPL	3-gramy WER [%]	4-gramy PPL	4-gramy WER [%]
referencia								
25k	25 025	12.12	472.52	29.80	368.14	28.48	362.49	28.48
50k	50 107	6.85	566.23	20.75	435.74	19.60	429.67	19.59
75k	75 271	4.82	615.25	17.64	471.37	16.51	464.83	16.58
100k	100 558	3.74	644.95	16.06	493.63	14.83	486.74	14.88
125k	125 930	3.05	669.48	15.05	512.10	13.94	505.00	13.94
navrhnutá segmentácia v tvare „kmeň-koncovka“								
25k	25 007	4.43	199.10	27.04	119.83	24.56	112.85	24.33
50k	50 234	2.36	211.67	22.95	124.99	20.48	117.60	20.41
75k	75 506	1.73	217.35	21.87	127.81	19.45	120.26	19.24
100k	100 132	1.47	220.24	21.28	129.51	18.73	121.90	18.60
116k	116 738	1.38	221.28	21.17	130.11	18.90	122.46	18.61
navrhnutá segmentácia v tvare „kmeň-koncovka“ - upravený slovník								
25k	25 011	4.64	198.95	27.58	120.79	25.29	113.25	25.09
50k	50 070	2.40	211.60	23.04	125.08	20.55	117.68	20.42
75k	75 196	1.73	217.45	21.79	127.90	19.36	120.34	19.10
100k	100 262	1.47	220.25	21.34	129.51	18.82	121.89	18.66
116k	116 738	1.38	221.28	21.17	130.11	18.90	122.46	18.61
Morfessor - segmentácia v tvare „kmeň-koncovka“								
25k	25 006	7.49	323.25	27.04	219.08	25.24	211.97	25.16
50k	50 162	3.87	365.35	21.10	244.20	19.16	236.04	19.27
75k	75 033	2.54	389.92	19.31	255.47	17.47	246.84	17.45
100k	100 352	1.88	396.99	18.45	263.69	16.56	254.82	16.55
125k	126 969	1.47	404.47	17.81	268.42	16.01	259.38	16.06
Morfessor rozšírený o navrhnutú segmentáciu v tvare „kmeň-koncovka“								
25k	25 011	4.28	198.13	27.67	120.40	24.94	113.46	24.77
50k	50 139	2.10	210.04	23.71	125.67	21.11	118.26	21.01
75k	75 033	1.40	215.60	22.80	128.49	20.15	120.89	19.92
100k	101 621	1.02	218.96	22.57	130.38	19.93	122.60	19.72
125k	125 583	0.83	221.19	22.76	131.51	20.01	123.75	19.86
Sylseg - slabičná segmentácia								
25k	25 090	4.02	97.49	42.23	50.22	36.33	42.04	35.26
50k	50 759	3.78	99.44	41.96	51.23	36.29	42.89	35.15
62k	62 388	3.75	99.80	42.17	51.42	36.49	43.06	35.23

modelov vzhľadom k referenčným celoslovným modelom jazyka. Štatistiku počtu najviac vyskytovaných morf a slov získaných v jednotlivých slovníkoch získaných z trénovacieho korpusu a počtu n-gramov v jazykových modeloch je možné vidieť v Tabuľke 1 a hodnoty počtu OOV tvarov, perplexity modelu a chybovosti systému LVCSR v Tabuľke 2.

Čo sa týka samotných výsledkov, z pohľadu delenia slov na čo najväčší počet subslovných jednotiek s ohľadom na ich prekryv, sme najlepšie hodnoty dosiahli pri nami navrhnutom princípe segmentácie slov na kmene a koncovky, aj keď segmentáciou pomocou nástroja Morfessor sme vyprodukovali väčší počet unikátnych koncoviek, ktoré sa však asi v tretine prípadov z väčšej časti prekrývali. Daný spôsob segmentácie nám však vyprodukoval aj najväčší počet n-gramov vo vytvorených jazykových modeloch. Naopak, navrhnutým spôsobom delenia slov aj v spojení so segmentáciou pomocou nástroja Morfessor sme dosiahli výraznú redukciu počtu mimoslovníkových slov približne o dve tretiny, už aj v prípade menších slovníkov. To sa odrazilo aj v hodnotách perplexity modelov, ktoré prakticky vo všetkých prípadoch dosahovali najnižšie hodnoty, a tak ako v prípade počtu OOV, aj v tomto prípade ich hodnota klesla približne na tretinu. K zhoršeniu došlo len v prípade chybovosti systému LVCSR, ktorá sa v priemere zvýšila približne o 3.92% WER v prípade bigramových modelov, o 1.75% WER pri trigramoch a o 1.54% WER pri štvorgramových modeloch, v absolútnej miere. K zlepšeniu hodnotiacich mier nedošlo ani úpravou slovníka a zavedením všetkých možných koncoviek do jednotlivých slovníkov, ako je to možné vidieť v tretej časti Tabuľky 2. Najlepšie výsledky presnosti rozpoznávania pomocou morfémových modelov sme však dosiahli pri upravenej segmentácii nástroja Morfessor. Vo všetkých doposiaľ spomenutých prípadoch sme však iba pri veľkosti slovníka $25k$ dosiahli výrazné zníženie miery výskytu chýb v rozsahu od 8.20% do 12.79% WER, relatívne. Naopak, aj pri najnižších hodnotách perplexity modelov, výsledky slabičnej segmentácie pomocou nástroja Sylseg produkovali najvyššie hodnoty WER, z čoho možno usúdiť, že tento typ segmentácie nie je príliš vhodný pri modelovaní slovenského jazyka pomocou morfémových modelov.

Na záver je vhodné ešte poznamenať, že v prípade modelovania slovenského jazyka pomocou modelov založených na kmeňoch a koncovkách slov sú trigramové modely úplne postačujúce. Pri použití modelov vyššieho rádu nebol pozorovaný výrazný posun hodnôt perplexity modelu ani chybovosti systému LVCSR. Ako najvhodnejší typ segmentácie sa tak javí nami navrhnutý

spôsob, príp. modifikovaná verzia segmentácie slov pomocou nástroja Morfessor, ktorého účinnosť by sme mohli v budúcnosti zvýšiť zavedením *modelovo-založeného prístupu segmentácie*, ktorý nám tento nástroj poskytuje, narozdiel od segmetácie obmedzenej slovníkom. Avšak z pohľadu voľby najvhodnejšej veľkosti slovníka v súčasnosti nevieme presne povedať, aká veľkosť slovníka by mala byť postačujúca pri návrhu efektívnych morfémových modelov slovenského jazyka pre pamäťovo-nenáročné aplikácie systému LVCSR v konkrétnych úlohách.

VI. Záver

Tento článok bol zameraný na analýzu metód automatickej segmentácie slov v slovenčine a na štatistické modelovanie jazyka pomocou morfémových modelov. Keďže slovenčina patrí k skupine vysoko-flektívnych jazykov, u ktorých flexia je sústredená na konci slova, rozhodli sme problematiku segmentácie slov zúžiť len v rámci delenia na kmeň a koncovku slova. Vzhľadom na to, že doposiaľ navrhnuté metódy segmentácie slov v slovenčine nezvažovali podmienky spĺňajúce delenie slov na morfo-fonologických hraniciach, vytvorili sme vlastný spôsob segmentácie slov na kmene a koncovky, ktorý v porovnaní s inými prístupmi dosiahol najlepšie výsledky počtu mimoslovníkových slov a hodnôt perplexity modelu s miernym zhoršením chybovosti systému LVCSR, v úlohe automatického prepisu spravodajských relácií v slovenčine. Ďalší výskum by mal byť zameraný na zdokonalenie dátami-riadených algoritmov segmentácie slov bez dohľadu, napr. vhodnou modifikáciou nástroja Morfessor, s cieľom zvýšiť kvalitu a znížiť pamäťové nároky morfémových modelov slovenského jazyka v systéme LVCSR.

Poďakovanie

Táto práca bola vytvorená realizáciou projektu Výskum modulov pre inteligentné robotické systémy (kód ITMS projektu: 26220220141) na základe podpory operačného programu Výskum a vývoj, financovaného z Európskeho fondu regionálneho rozvoja (100%).

Literatúra

[1] P. Majewski, *Syllable Based Language Model for Large Vocabulary Continuous Speech Recognition of Polish*, In: Proc. of the 11^{th} International Conference on Text, Speech and Dialogue, TSD'2008, Brno, Czech Republic, pp. 397–401, 2008.

[2] M. Creutz, T. Hirsimäki, M. Kurimo, A. Puurula, J. Pylkkönen, V. Siivola, M. Varjokallio, E. Arisoy, M. Saraclar and A. Stolcke, *Analysis of Morph-Based Speech Recognition and the Modeling of Out-of-Vocabulary Words across Languages*, In: NAACL-HLT 2007, pp. 380–387, 2007.

[3] W. J. Byrne, J. Hajič, P. Krbec, P. Ircing and J. Psutka, *Morpheme Based Language Models for Speech Recognition of Czech*, In: Proc. of the 3^{rd} International Workshop on Text, Speech Dialogue, TSD'2000, Brno, Czech Republic, pp. 211–216, 2000.

[4] T. Rotovnik, M. S. Maučec and Z. Kačič, *Large Vocabulary Continuous Speech Recognition of an Inflected Language using Stems and Endings*, In: Speech Communication, Vol. 49, No. 6, pp. 437–452, 2007.

[5] A. Karpov, I. Kipyatkova and A. Ronzhin, *Very Large Vocabulary ASR for Spoken Russian with Syntactic and Morphemic Analysis*, In: Proc. of INTERSPEECH'2011, Florence, Italy, pp. 3161–3164, 2011.

[6] P. Shone and D. Jurafsky, *Knowledge-free Induction of Morphology using Latent Semantic Analysis*, In: Proc. of the 2^{nd} Workshop on Learning Language in Logic and the 4^{th} Conference on Computational Natural Language Learning, Lisbon, Portugal, pp. 67–72, 2000.

[7] Z. S. Harris, *From Morpheme to Phoneme*, In: Language, Vol. 31, No. 2, pp. 190–222, 1955.

[8] S. Bordag, *Unsupervised Knowledge-Free Morpheme Boundary Detection*, In: Proc. of International Conference on Recent Advances in Natural Langauge Processing, RANLP-2007, Borovets, Bulgaria, pp. 1–7, 2005.

[9] A. Barron, J. Rissanen and B. Yu, *The Minimum Description Length Principle in Coding and Modeling*, In: IEEE Transaction on Information Theory, Vol. 44, No. 6, pp. 2743–2760, 1998.

[10] M. Creutz and K. Lagus, *Unsupervised Morpheme Segmentation and Morphology Induction from Text Corpora using Morfessor 1.0*, In: Computer and Information Science, Vol. Report A81, pp. 1–27, 2005.

[11] T. Hirsimäki, J. Pylkkönen and M. Kurimo, *Importance of Higher-Order N-Gram Models in Morph-Based Speech Recognition*, In: IEEE Transaction on Audio, Speech, and Language Processing, Vol. 17, No. 4, pp. 724–732, 2009.

[12] J. Ivanecký, *Štatistický prístup pri určovaní slabičných hraníc*, In: Proc. of SLOVKO'2003, Bratislava, Slovak Republic, s. 11, (in Slovak), 2003.

[13] M. Mirilovič and J. Juhár, *Morphological Segmentation of Word Units for Large Vocabulary Automatic Speech Recognition in Slovak*, In: Proc. of the 3^{rd} Baltic Conference on Human Language Technologies, Kaunas, Lithuania, pp. 189–196, 2007.

[14] D. Zlacký, *Morfémový model slovenského jazyka v systémoch automatického rozpoznávania reči*, Diplomová práca, Katedra elektroniky a multimediálnych telekomunikácií, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika, s. 54, 2012.

[15] D. Hládek, J. Staš and J. Juhár, *Word Clustering for a Slovak Class-Based Language Model*, In: Journal of Electrical and Electronics Engineering, Vol. 5, No. 1, pp. 1–4, (to be published), 2012.

[16] JUĽŠ-SNK, *Ručne morfologicky anotovaný korpus*, Jazykovedný ústav Ľudovíta Štúra, Slovenský národný korpus, Bratislava, Slovenská republika, 2007. URL: http://korpus.juls.savba.sk/stats.html

[17] A. Stolcke, *SRILM – An Extensible Language Modeling Toolkit*, In: Proc. of ICSLP'2002, Denver, Colorado, USA, pp. 901–904, 2002.

[18] D. Hládek and J. Staš, *Text Mining and Processing for Corpora Creation in Slovak Language*, In: Journal of Computer Science and Control Systems, Vol. 3, No. 1, pp. 65–68, ISSN 1844-6043, 2010.

[19] S. Darjaa, M. Ceňak, M. Trnka and M. Rusko, *Effective Triphone Mapping for Acoustic Modeling in Speech Recognition*, In: Proc. of INTERSPEECH'2011, Florence, Italy, pp. 1717–1720, 2011.

[20] T. Lee, T. Kawahara and K. Shikano, *Julius – An Open Source Real-time Large Vocabulary Recognition Engine*, In: Proc. of EUROSPEECH'2001, Aalborg, Denmark, pp. 1691–1694, 2001.

Krátky vstup do problematiky analýzy chýb v systémoch rozpoznávania reči v slovenčine

Ján STAŠ, Pavol KOŠČ, Jozef JUHÁR

Katedra elektroniky a multimediálnych telekomunikácií,
Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach,
Park Komenského 13, 042 00 Košice, Slovenská republika

jan.stas@tuke.sk, pavol.kosc@student.tuke.sk, jozef.juhar@tuke.sk

Abstract **– Pri návrhu akéhokoľvek systému komunikujúceho s človekom prirodzenou rečou je nutné brať do úvahy rôzne činitele vplývajúce na výslednú úspešnosť rozpoznania rečového prejavu a na robustnosť systému ako celku. Tieto činitele sú dané najmä vonkajšími vplymi prostredia či odlišným spôsobom artikulácie rečníkov a priamo vplývajú na úspešnosť rozpoznávania. Pri štatistickom rozpoznávaní reči je však nutné upriečť pozornosť aj na chyby spôsobené samostatnými komponentami, z ktorých daný systém pozostáva. Nedostatok znalostných zdrojov v procese trénovania modelov jazyka a reči, ale aj zložitosť úlohy rozpoznávania, najmä v prípade jazykov s bohatou morfológiou a neusporiadanosťou jazyka, zvyšuje riziko chybovosti a vplýva aj na celkovú robustnosť systému. Z toho dôvodu je vhodné upriečť pozornosť na metódy eliminácie chybovosti pomocou diagnostiky, ktorá odhaľuje príčiny vzniku chýb, identifikuje chybný komponent a pomocou vhodných metód analyzuje vybrané druhy chýb, odstranením ktorých v ďalšom kroku rozpoznávania môžeme efektívne redukovať ich množstvo vnesené pri návrhu konkrétneho systému. Tento článok podáva krátky prehľad o taxonómii spracovania chýb, zaoberá sa kategorizáciou a analýzou vybraných druhov chýb, hodnotiacimi mierami a spôsobmi eliminácie chybovosti v systémoch rozpoznávania plynulej reči v slovenskom jazyku.**

Kľúčové slová **– analýza chýb, identifikácia chýb, kategorizácia chýb, rozpoznávanie plynulej reči s veľkým slovníkom**

I. Úvod

Analýza chýb je neoddeliteľnou súčasťou zvyšovania kvality a robustnosti systémov na automatické rozpoznávanie reči (ARR). Presnosť týchto systémov závisí najmä od správnej artikulácie rečníkov využívajúcich aplikáciu na rozpoznávanie reči, je daná tiež akustickým prostredím, v ktorom sa rečník nachádza, zložitosťou úlohy rozpoznávania, to znamená, či sa jedná o rozpoznávanie izolovaných slov, o spojitý diktát alebo o hovorenú resp. spontánnu reč, no vo veľkej miere aj samostatnými komponentami, z ktorých daný systém pozostáva, t. j. akustickým a jazykovým modelom, slovníkom a dekodérom [1]. Obzvlášť je to v prípade jazykov s bohatou morfológiou a najmä takých, ktoré trpia nedostatkom textových zdrojov a rečových databáz, medzi ktoré patrí aj slovenčina, keďže robustnosť akustického a jazykového modelu vychádza z dát, na ktorých boli modely natrénované a priamo vplýva na výsledok rozpoznávania.

Keďže nie je možné všetky chyby v systéme analyzovať ručne, hľadajú sa preto rôzne spôsoby automatickej diagnostiky [2]. Táto úloha je pomerne náročná a nie všetky chyby vieme v systéme ARR dôsledne identifikovať, analyzovať, kategorizovať a odstrániť. Pri automatickej diagnostike sa v súčasnosti často ako klasifikátory využívajú metódy založené na rozhodovacích stromoch (binárne, regresné, CART a pod.), podrobne opísané v [3].

Čo sa týka samotného problému analýzy chýb, najprv je nutné chyby v rozpoznanom texte správne automaticky identifikovať, analyzovať príčinu a zdroj chýb z pohľadu jednotlivých komponentov systému, to znamená, či sa jedná o chyby spôsobené nevhodným predspracovaním vstupného signálu, slabým pokrytím slovnej zásoby, nedostatočnosťou trénovacích dát pri tvorbe akustických a jazykových modelov, alebo chybami v prehľadávacej stratégii počas rozpoznávania. Následne môžeme pristúpiť k ich odstráneniu buď zásahom do jednotlivých komponentov systému alebo k čiastočnej eliminácii korekciou v kroku postspracovania textových dát po rozpoznaní.

Väčšinu chýb vieme v súčasnosti eliminovať už pri samotnom návrhu systému ARR, a to použitím vhodných metód parametrizácie rečového signálu vstupujúceho do procesu rozpoznávania reči, výberom čo možno najviac frekventovaných slov pri zostavovaní slovníka, generovaním alternatívnych výslovností pri fonetickom prepise slov, inšpirovať sa možno tiež najnovšími postupmi pri tvorbe akustických a jazykových modelov s ohľadom na štruktúru daného jazyka, ďalej adaptáciou akustického modelu na rečníka, resp. jazykového modelu vzhľadom na oblasť, v ktorej sa rozpoznávanie uskutočňuje, použitím efektívnych algoritmov pri návrhu dekodéra reči, resp. viacprechodového dekódovania, či dôsledným postspracovaním textových dát po rozpoznaní.

Tento článok podáva krátky vstup do problematiky analýzy chýb v systémoch ARR pre slovenský jazyk. V nasledujúcich častiach budú v krátkosti zhrnuté jednotlivé mechanizmy spracovania chýb pomocou rozšírenej taxonómie, navrhnutej primárne pre systémy ARR. Následne v rámci kategorizácie predstavíme jednotlivé typy chýb, ktoré sa v hypotézach daných automatickým prepisom rečového prejavu objavujú. Túto kategorizáciu rozšírime aj o prípady zahrňujúce okolitý kontext slov. V ďalšej kapitole zameranej na analýzu chýb v slovenčine opíšeme vybrané druhy chýb, ktoré sme reálne v hypotézach po rozpoznaní spozorovali, zdôvodníme ich pôvod a predstavíme spôsoby ich eliminácie. Predposledná kapitola podáva prehľad o mierach v súčasnosti najviac používaných pri hodnotení systémov na automatické rozpoznávanie reči. Na záver budú spomenuté hlavné prínosy článku s ďalším smerovaním v oblasti identifikácie, analýzy a redukcie chýb v systémoch na automatické rozpoznávanie reči v slovenčine.

II. Taxonómia spracovania chýb

Pri vývoji systémov na automatické rozpoznávanie reči v reálnych aplikáciách, kde sa nie je možné vyhnúť vneseným chybám, je nutné sa zamerať aj na problematiku spracovania chýb. Táto časť poukazuje na nové perspektívne mechanizmy spracovania chýb. Navrhovaná taxonómia, primárne z oblasti rozpoznávania reči a príbuzných komunikačných odborov, prispieva ku štandardizácii spracovania chýb v rečových aplikáciách. V rámci tejto taxonómie bolo navrhnutých päť základných mechanizmov spracovania chýb [3]:

1) **prevencia chýb** - skúmajúca predpoklad o výskyte chýb s cieľom *redukovať chybovosť* systému ARR zahrnutím tejto informácie v následnom pokuse (pri viacprechodovom rozpoznávaní reči, resp. pri ďalšom spustení systému). Rozlišujeme:
   a) ***redukciu chýb návrhom*** - v „*off-line*“ aplikácii použitím metód diskriminatívneho trénovania akustického modelu, modelovaním jazyka s obmedzeným slovníkom, doménovo-orientovaným modelovaním jazyka [4], použitím gramatiky zahrňujúcej širší kontext slov a pod.;
   b) ***redukciu chýb kontextom*** - v „*on-line*“ aplikácii použitím n-gramov vyššieho rádu v jazykovom modeli, príp. rôznych druhov jazykových modelov pri viacprechodovom rozpoznávaní reči, modelovaním reči pomocou trifónových alebo quadrofónových skrytých Markovovych modelov (v angl. „*hidden Markov model*“, skr. HMM), príp. adaptáciou akustického modelu na rečníka [5] alebo jazykového modelu na prebiehajúcu tému [4];
2) **detekcia chýb** - orientovaná na *identifikáciu príznakov* a *miest vzniku chýb*, napr. pomocou dátami-riadených prístupov využívajúcich čiastkové skóre akustického a jazykového modelu v porovnaní s celkovým skóre z dekódera [2];
3) **korekcia chýb** - pomocou *manuálnej* alebo *automatickej úpravy* (postspracovania) textových dát po rozpoznaní, napr. vyradením nevhodných kandidátov z „N-best“ zoznamov pomocou sémantických modelov počas rozpoznávania [6], preddefinovanou gramatikou zahrňujúcu širší kontext slov, využitím morfologických analyzátorov (napr. kontrolou zhodnosti gramatických kategórií prídavného mena s podstatným menom) a pod.;
4) **predikcia chýb** - hľadajúca odpovede na otázky typu: „*kedy*“ najčastejšie chyby vznikajú a o „*aké*“ chyby, resp. zdroje chýb sa jedná;
5) **diagnostika chýb** - zaoberajúca sa problémami typu: „*ako vplývajú*“ jednotlivé časti systému ARR na výslednú chybovosť a „*čo je príčinou*“ tvorby týchto chýb [2]. Rozlišujeme:
   a) ***diagnostiku intrinzických chýb***, medzi ktoré radíme:
      - *chyby slovníka* - spôsobené vysokým počtom mimoslovníkových tvarov slov (v angl. „*out-of-vocabulary*“, skr. OOV), chýbajúcimi alternatívnymi výslovnosťami v slovníku s fonetickým prepisom, príp. nejednoznačnosťou pri rozpoznávaní zložených či morfologicky-zložitých alebo cudzích slov;
      - *chyby akustického modelu* - týkajúce sa preferencie vnútroslovných a medzislovných trifón, najmä v prípade reprezentácie nepozorovaných trifón (napr. nevhodným spôsobom ich mapovania);

Tabuľka 1
Príklady chybných segmentov

druhy chýb	referencia	hypotéza
chyby zámenou	*dúha *pohltil *pristihnúť	*túha *pohltila *pristihnul
chyby vložením	uplatnený *** *** aute *** dôvodov	uplatnený v a aute z dôvodov
chyby zmazaním	v súlade právo k o odmenách	*** súlade právo *** *** odmenách
chyby transpozíciou	*zdvihol *korán *marenie	*zdvihlo *konár *meranie
chyby delením	*zákazníkmi *** *juhomaďarského *** *Drutarovských ***	*zákazník mi *juh maďarského *Drutarovský ich
chyby zreťazením	*je prístupná *nad Váhom *tri krát	*neprístupná *** *nadváhou *** *trikrát ***

- *chyby jazykového modelu* - vyvolané ústupovou schémou vyhladzovania pri javoch nepozorovaných v trénovacích dátach (preferencia bigramovej pravdepodobnosti pred trigramovou s veľmi malou pravdepodobnosťou výskytu);
- *chyby v prehľadávacej stratégii* - zlým vyhodnotením výsledného skóre z dekodéra reči (najpravdepodobnejšej hypotézy z čiastkového skóre určeného akustickým a jazykovým modelom);

b) ***diagnostiku extrinzických chýb***, ku ktorým patria:

- *chyby rečníka* - vnesením rôznych mimorečových prejavov ako je zaváhanie či zakašlanie, príp. iných dysfluentných javov spôsobených rečníkom, t. j. chyby spôsobené nesprávnou artikuláciou, výslovnosťou a pod.;
- *chyby okolia* - spôsobené akustickým prostredím (prítomnosťou okolitého šumu a rušenia z prenosového kanála).

III. Kategorizácia chýb

V rámci kategorizácie štandardne rozdeľujeme chyby do troch základných skupín, a to na:

- **chyby zámenou** S (v angl. „*substitution errors*“) - kedy slovo v referencii bolo zamenené v hypotéze za iné (foneticky príbuzné) slovo;
- **chyby vložením** I (v angl. „*insertion errors*“, alebo tiež „*false alarms*“) - v prípade, že v hypotéze sa objavilo slovo, ktoré sa v referencii vôbec nevyskytuje;
- **chyby zmazaním** D (v angl. „*deletion errors*“, alebo tiež „*false rejections*“) - kedy slovo v referencii bolo z hypotézy vymazané.

Špeciálnym prípadom chýb zámenou sú tzv. **chyby transpozíciou** T (v angl. „*transposition*“ alebo „*reversal errors*“), kedy dochádza k chybnej zámene dvoch alebo viacerých tých istých znakov v rámci jedného slova.

Chyby zámenou môžeme tiež v závislosti od okolitého kontextu ďalej deliť na:

- **chyby delením** (v angl. „*split errors*“) - kedy v hypotéze po rozpoznaní dochádza k rozdeleniu jedného slova na dve alebo viaceré segmenty (chybné, foneticky príbuzné slová). Častokrát je to prípad zložených slov, a to najmä zložených čísloviek, resp. prípad zlého rozpoznania predpony alebo prípony slova, ktorá je od slova oddelená a zamenená za iné (často krátke, jednoslabičné, foneticky príbuzné) slovo. Dochádza tu k preferencii medzislovných trifón pred vnútroslovnými, ktoré sa v trénovacej množine pri tvorbe akustického modelu vyskytovali menej pravdepodobne.
- **chyby zreťazením** (v angl. „*joint errors*“) - kedy v hypotéze po rozpoznaní dochádza k zreťazeniu priľahlých dvoch alebo viacerých slov do jedného (v mnohých prípadoch aj foneticky príbuzného) slova. Tento jav je opačný k chybám delením, kedy pri vyhodnocovaní skóre dekodérom z akustického modelu v hypotéze dochádza k preferencii vnútroslovných trífón pred medzislovnými.

Príklady spomenutých druhov chýb s označením jednotlivých chybných segmentov, kde znak * označuje chybné slovo a znaky *** slovo nenachádzajúce sa v referencii alebo v hypotéze, sú znázornené v Tabuľke 1.

IV. Analýza chýb v slovenčine

Z najbežnejších chýb, ktoré sme zistili pri analýze textových dát po rozpoznaní systémom na automatické rozpoznávanie plynulej reči v slovenčine, možno spomenúť [7]:

1) **správne rozpoznanie slovného kmeňa, no chybné rozpoznanie koncovky** - reprezentované chybami zámenou, tiež označované ako *chyby sprava*, vyplývajúce z flexie slovenského jazyka. V tomto prípade často dochádza k zámene znakov vnútri slova, napr. znaku „*-m*“ za „*-b*“, príp. „*-n*“ za „*-d*“, alebo k zámene koncovky, resp. prípony slova s foneticky veľmi podobnou výslovnosťou, napr. v prípade derivačnej morfémy „*-ú*“, ktorá je často zamenená za morfému „*-út*“, napr. u slov „*pristihnú*“ a „*pristihnút*“, resp. vzájomnou zámenou koncoviek „*-l*, *-la*, *-li*, *-lo*, príp. *-ol*“, napr. v slove „*spadol*“, ktoré bolo nesprávne zamenené za slovo „*spadlo*“. Tieto chyby sú často spôsobené zlou výslovnosťou rečníka, rovnako tiež akustickým a jazykovým modelom, príp. zlým vyhodnotením skóre v procese dekódovania vyslovenej postupnosti slov.
2) **rozdelenie slova na dve alebo viacero častí** - jedná sa o tzv. problém zložených slov (najmä zložených čísloviek), ale tiež viacslabičných slov, kedy dochádza k preferencii medzislovných trifón pred vnútroslovnými v reprezentácii reči pomocou akustického modelu a deleniu slova na viacero častí. Tieto chyby môžu byť taktiež spôsobené slovníkom a jazykovým modelom, v ktorých sa okrem štandardných slov môžu vyskytovať aj zložené tvary, ako napr. slová: „*lásky-plný*“, „*auto-pilot*“, či „*juho-americký*“, kde znak „-“ určuje hranicu delenia. V určitých prípadoch dochádza aj k zámene niektorého z dvojice, resp. n-tice slov vytvorených chybnou segmentáciou referencie delením. Potom rozlišujeme:
   - *chyby delením zľava* - v prípade oddelenia predpony slova a jej zámenou za foneticky príbuzné slovo. Častokrát sa jedná o prípad záporných slov, napr. slovo „*neznáma*“ zamenené za dvojicu slov „*mne známa*“ alebo „*nešťastný*“ za „*je šťastný*“;
   - *chyby delením sprava* - kedy dochádza k oddeleniu koncovky slova s jej následnou zámenou za iné, foneticky príbuzné slovo, napr. slovo „*podnikateľmi*“ zamenené za dvojicu slov „*podnikateľ mi*“ alebo „*pokyvkáva*“ za „*pokým káva*“.
3) **zreťazenie dvoch alebo viacerych slov do jedného** - označovanie tiež ako chyby zreťazením, pri ktorých dochádza k preferencii vnútroslovných trifón pred medzislovnými v akustickej reprezentácii modelom reči. K týmto chybám možno zaradiť aj chyby spôsobené zamietnutím krátkych jednoslabičných slov v spojení s dlhými viacslabičnými slovami, ktoré sú často lepšie rozpoznateľné, vyskytujúce sa najmä na začiatku rečového prejavu alebo v miestach, kde rečník zaváha či zakašle. Na elimináciu takéhoto druhu chýb a na zlepšenie presnosti rozpoznávania sa v systémoch rozpoznávania reči potom často využívajú aj jazykové modely s viacslovnými výrazmi, ako je to ukázané v [8]. Obdobne ako v predchádzajúcom odseku rozlišujeme:
   - *chyby zreťazením zľava* - v prípade zlúčenia krátkeho jednoslabičného slova s dlhým viacslabičným slovom, ktoré za ním nasleduje, t. j. zlúčenie predložky alebo spojky s nasledujúcim slovom, napr. dvojica slov „*od dlžníka*“ zlúčená do „*poddlžníka*“, alebo pri zlúčení dvoch slov do jedného zloženého tvaru, napr. slová „*juh balkánskeho*“ reprezentované slovom „*juhobalkánskeho*“;
   - *chyby zreťazením sprava* - prípad veľmi podobný chybám delením sprava, kedy dochádza naopak k zlúčeniu dvoch foneticky príbuzných slov do jedného, napr. u slov „*kraj sa*“, ktoré boli nesprávne zlúčené do slova „*krajca*“ alebo zamietnutie v poradí druhého slova v bigrame obsahujúcom spojku „*cena a*“ za slovo „*cenná*“.
4) **problém správnej výslovnosti neštandardných slov** - menej častý, avšak veľmi dôležitý problém správnej výslovnosti skratiek, značiek a akronymov v slovníku s fonetickým prepisom, ktoré môžu byť vyslovované buď jedným prúdom reči, prípad tzv. splývavej výslovnosti, napr. slovo „*SĽUK*“ ako [`s l u k`], alebo hláskované po slabikách, napr. „*KDH*“ ako [`k a: d e: h a:`], resp. s vloženou švou, napr. „*ŽSR*“ ako [`ž ə s ə r ə`]. Výslovnosť môže byť tiež kombinovaná, napr. „*KFOR*“ ako [`k a: f o r`], príp. sa môže jednať o cudzokrajinné skratky, čiastočne zdomácnené, napr. „*OECD*“ vyslovovaná ako [`o i: s i: d i:`]. Rovnaký problém nastáva aj v prípade výslovnosti cudzokrajinných slov a odborných termínov v cudzom jazyku, ktoré je možné zlúčiť aj do viacslovných výrazov a v slovníku výslovnosti, resp. v modeli jazyka, reprezentovaného formou jedného slova, napr. latinský výraz „*de facto*“ foneticky prepísaný ako [`d e f a k t o`] a modelovaný v jazyku ako [`de_facto`]. Problém správnej výslovnosti potom zahŕňa aj viacero alternatív, ktoré je nutné v slovníku výslovnosti správne, často manuálne, prepísať.
5) **výskyt slov z okraja slovnej zásoby** - tieto slová sú často dekodérom v procese rozpoznávania zle vyhodnotené. Vyplýva to z ich nízkeho frekvenčného výskytu v trénovacích dátach,

čo v jazykovom modeli odráža jav veľmi málo pravdepodobný na to, aby bol lingvistickým dekodérom správne rozpoznaný. Ako príklad možno uviesť slová „*ikonografický*“ či „*rozhadzovačnosťou*“ pri doménovo-orientovanom rozpoznávaní reči z oblasti súdnictva.

6) **iné, menej systematické chyby**, ktoré zahŕňajú skomoleniny spôsobené nevhodnou artikuláciou, resp. výslovnosťou rečníka, ktoré dekodér často vyhodnotí ako zhluk krátkych jednoslabičných alebo hláskovaných slov, napr. slovo „*krízových situáciách*“ v referencii bolo v hypotéze nesprávne rozpoznané ako „*kríza ich C dva C tri L*“.

Odhad miesta výskytu chýb v prípade chýb zámenou, transpozíciou, delením alebo zreťazením sa zvyčajne určuje výpočtom *Levenshteinovej vzdialenosti* (v angl. „*Levenshtein*“ alebo „*edit distance*“), ktorá vychádza z porovnávania grafém obsiahnutých v referenčnom slove a slove, resp. slovách v hypotéze po automatickom prepise. Typicky sa uvažuje s prekryvom slov v 70-tich percentách a viac, aby bolo možné jednoznačne preukázať, že došlo k zámene prefixálnych alebo sufixálnych morfém so zachovaním slovného kmeňa v prípade foneticky príbuzných slov (chyby akustického alebo jazykového modelu) a nie k zámene slova za iné, foneticky menej pravdepodobné slovo (čo indikuje chyby v prehľadávacej stratégii).

Na záver tejto časti je vhodné spomenúť, že k eliminácii chybovosti a zlepšenia robustnosti systému na rozpoznávanie plynulej reči je možné tiež prispieť modelovaním vložených páuz v akustickom a jazykovom modeli či zavedením vybraných druhov modelov dysfluentných javov ako sú repetície alebo prolongácie štatisticky priamo do modelu jazyka [9].

V. Miery hodnotenia kvality systému ARR

Táto časť podáva prehľad o základných mierach hodnotenia systémov ARR, ktoré umožňujú dôkladnejšie analyzovať výkonnosť systému, poskytujú jasnejší výklad, najmä pokiaľ ide o ich použiteľnosť v koncových aplikáciách, hodnotia systém *priamo* a *objektívne* (automatickým spôsobom) a sú ľahko *interpretovateľné* a *modulárne* [10].

V súčasnosti najbežnejšie používanou mierou hodnotenia systému ARR je **miera chybovosti**, resp. **miera výskytu chýb** (v angl. „*word error rate*“, skr. WER), ktorá vychádza z *Levenshteinovej vzdialenosti* a je definovaná ako minimálna vzdialenosť (vážený súčet) medzi referenčnou sekvenciou slov a automatickým prepisom po rozpoznaní podľa vzťahu:

$$WER = \frac{S + D + I}{N_R}, \tag{1}$$

kde N_R je celkový počet slov v referencii, S odráža počet vložení v rozpoznanej hypotéze, D udáva počet slov zmazaných z hypotézy vzhľadom k referenčnému textu a I je počet slov vložených systémom ARR počas rozpoznávania, nevyskytujúcich sa v referencii. Okrem štandardnej miery výskytu chýb sa na hodnotenie používajú tiež miery ako:

- SER (v angl. „*sentence (utterance) error rate*“) - zovšobecnená na hodnotenie celkovej sekvencie slov obsiahnutých vo vete;
- CER, resp. LER (v angl. „*character (letter) error rate*“) - používaná pri hodnotení systémov pracujúcich so subslovnými jednotkami reči obsiahnutých v jazykovom modeli;
- PER (v angl. „*phoneme error rate*“) - v prípade rozpoznávania izolovaných foném (často bez použitia jazykového modelu).

Doplnkovou mierou k miere výskytu chýb je **miera úspešnosti** rozpoznania (v angl. „*word recognition rate*“, skr. WRR, alebo tiež „*accuracy*“, skr. Acc) definovaná vzťahom:

$$WRR\ [Acc] = 1 - WER = \frac{H - I}{N_R}, \tag{2}$$

kde $H = N_R - (S + D)$ udáva celkový počet slov rozpoznaných systémom ARR.

V prípade systémov na rozpoznávanie izolovaných slov sa častejšie používa **miera korektnosti** (v angl. „*word correct rate*“, skr. WCR, alebo tiež „*correctness*“, skr. $Corr$), ktorá oproti predchádzajúcim mieram nezvažuje chyby vložením I a je definovaná ako:

$$WCR\ [Corr] = \frac{H}{N_R}. \tag{3}$$

Ak sa vrátime ku kritériám spomenutých v úvode tejto časti, je zrejmé, že WER je priamou a objektívnou mierou hodnotenia systému ARR. Avšak hlavným nedostatkom hodnotenia pomocou WER je skutočnosť, že v určitých prípadoch nie je jednoducho interpretovateľná (v zmysle určenia správneho počtu slov v referencii N_R, kvôli vneseným vloženiam I) a málokrát tiež modulárna (nie je prenositeľná v rámci jednotlivých komponentov systému).

Jednou z alternatívnych mier k miere výskytu chýb je **miera zachovania slovnej informácie** (v angl. „*word information preserved*“), ktorá vyjadruje aproximovanú hodnotu podielu informácie obsiahnutú v referencii vzhľadom k rozpoznanej sekvencii slov a je definovaná ako:

$$WIP = \frac{H^2}{(H+S+D)(H+S+I)}. \tag{4}$$

WIP odstraňuje nedostatky hodnotenia pomocou WER, ktoré sú spôsobené asymetriou vzhľadom na počet vložení I a zmazaní D. Je odvodená zo vzájomnej informácie medzi referenčným textom a hypotézou po rozpoznaní a odráža druhú mocninu geometrického priemeru *presnosti* (v angl. „*precision*“) a *odozvy* (v angl. „*recall*“) systému ARR [10].

Medzi ďalšie dôležité miery hodnotenia týkajúce sa viac-menej len jednotlivých súčastí systému ARR možno zaradiť tiež:

- **počet mimoslovníkových slov** (v angl. „*out-of-vocabulary*“, skr. OOV), miera týkajúca sa pokrytia slovnej zásoby v jazyku (čím je hodnota nižšia, tým je pokrytie jazyka lepšie);
- **zložitosť jazykového modelu** (v angl. „*model perplexity*“, skr. PPL), ktorá vychádza z krížovej entropie a je definovaná ako reciproká hodnota váženého (geometrického) priemeru pravdepodobnosti pridelenej jazykovým modelom pre každé slovo v testovacích dátach (čím je hodnota nižšia, tým je model presnejší);
- **časová odozva** systému ARR (v angl. „*real-time factor*“, skr. xRT), ktorá sa udáva ako pomer skutočného času trvania procesu rozpoznávania k reálnemu času rozpoznávaného javu (čím je hodnota nižšia, tým sa čas rozpoznávania blíži viac k reálnemu času vyslovovania sekvencie slov užívateľom systému).

VI. Záver

V tomto článku sme sa zamerali na analýzu chýb v systémoch automatického rozpoznávania reči v slovenčine. V rámci taxonómie sme predstavili základné mechanizmy spracovania chýb, ktoré sa v súčasnosti využívajú pri návrhu systémov na automatickú diagnostiku chýb v rozpoznávani reči. Štandardnú kategorizáciu chýb delených na chyby zámenou, vložením a zmazaním sme rozšírili o nové pojmy, a to na chyby delením a chyby zreťazením, zvažujúce aj okolitý kontext slov. Analýzou výstupných hypotéz z rozpoznávania sme načrtli základné problémy vyskytujúce sa pri automatickom prepise rečových nahrávok v slovenčine a rozšírili sme si tak poznatky, ktoré nám môžu výrazne pomôcť pri návrhu systému na automatickú diagnostiku chýb, na ktorom v súčasnosti pracujeme. Tento nástroj by mal v budúcnosti poslúžiť na lepšiu identifikáciu miest a analýzu príčin a zdrojov chýb s cieľom zlepšiť úspešnosť a robustnosť doposiaľ navrhnutých systémov na automatický prepis reči do textu v slovenčine.

Poďakovanie

Táto práca bola vytvorená realizáciou projektu Výskum modulov pre inteligentné robotické systémy (kód ITMS projektu: 26220220141) na základe podpory operačného programu Výskum a vývoj, financovaného z Európskeho fondu regionálneho rozvoja (100%).

Literatúra

[1] J. Psutka, L. Müller, J. Matoušek a V. Radová, *Mluvíme s počítačem česky*, Academia, Praha, s. 752, 2006.
[2] H. Nanjo, A. Lee, and T. Kawahara, *Automatic Diagnosis of Recognition Errors in Large Vocabulary Continuous Speech Recognition System*, In: Proc. of ICSLP'2000, Beijing, China, pp. 1027–1030, 2000.
[3] M. Cerňak, *Spracovanie chýb v rozpoznávaní reči*, In: J. Juhár a kol.: Rečové technológie v telekomunikačných a informačných systémoch, EQUILIBRIA, s.r.o., Košice, Slovensko, pp. 283–300, 2011.
[4] J. Staš, D. Hládek and J. Juhár, *Language Model Adaptation for Slovak LVCSR*, In: Proc. of the International Conference on Applied Electrical Engineering and Informatics, AEI'2010, Venice, Italy, pp. 101–106, 2010.
[5] M. Papco and J. Juhár, *Comparison of Acoustic Model Adaptation Methods and Adaptation Database Selection Approaches*, In: Journal of Electrical and Electronics Engineering, Vol. 3, No. 1, pp. 147–150, 2010.
[6] H. Stehouwer and M. van Zaanen, *Language Models for Contextual Error Detection and Correction*, In: Proc. of EACL 2009 Workshop on Computational Linguistic Aspects of Grammatical Inference, Athens, Greece, pp. 41–48, 2009.
[7] Š. Beňuš, M. Cerňak, M. Rusko, M. Trnka, S. Darjaa and R. Sabo, *Semi-Automatic Approach to ASR Errors Categorization in Multi-Speaker Corpora*, In: Proc. of the 6^{th} International Conference on Natural Language Processing, Multilinguality, SLOVKO'2011, Modra, Slovak Republic, pp. 9–17, 2011.
[8] J. Staš, D. Hládek, M. Trnka and J. Juhár, *Automatic Extraction of Multiword Expressions using Linguistic Constraints for Slovak LVCSR*, In: Proc. of the 6^{th} International Conference on Natural Language Processing, Multilinguality, SLOVKO'2011, Modra, Slovak Republic, pp. 138–145, 2011.
[9] A. Stolcke, E. Shriberg, R. Bates, M. Ostendorf, D. Hakkami, M. Plauche, G. Tür and Y. Lu, *Automatic Detection of Sentence Boundaries and Disfluencies based on Recognized Words*, In: Proc. of ICSLP'98, Sydney, Australia, pp. 0059, 1998.
[10] I. McCowan, D. Moore, J. Dines, D. Gatiza-Perez, M. Flynn, P. Wellner and H. Bourlard, *On the Use of Information Retrieval Measures for Speech Recognition Evaluation*, Tech. Rep.: IDIAP Research Report 04-73, Martigny, Switzerland, pp. 13, 2005.

Morfologicky-založené modely slovenského jazyka pre systémy rozpoznávania plynulej reči

Ján STAŠ, Tibor ROVENSKÝ, Jozef JUHÁR

Katedra elektroniky a multimediálnych telekomunikácií,
Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach,
Park Komenského 13, 042 00 Košice, Slovenská republika

jan.stas@tuke.sk, tibor.rovensky@student.tuke.sk, jozef.juhar@tuke.sk

Abstrakt **– Slovenský jazyk patrí do skupiny vysoko-flektívnych jazykov, ktoré sú charakteristické neusporiadanosťou jazyka, bohatou morfológiou, vzrastajúcim slovníkom a nedostatkom textových zdrojov. V systémoch rozpoznávania reči a modelovaním jazyka pomocou n-gramových modelov táto skutočnosť vytvára problém vnútorného rozptylu trénovacích dát. Na elimináciu daného problému je často nutné upriať pozornosť na morfologicky-založené modely použitím morfologických pravidiel obsiahnutých priamo v jazyku a modelovať jazyk pomocou slovných tried. Keďže morfológia slovenského jazyka je sústredená na konci slova, navrhli sme preto pravidlá na automatickú segmentáciu kmeňov a koncoviek slov a tento prístup sme použili aj pri zhlukovaní slov do slovných tried. Táto myšlienka bola neskôr aplikovaná aj pri tvorbe štatistických modelov slovenského jazyka pomocou slovných tried odvodených od kmeňov a koncoviek slov. Navrhnutý spôsob zhlukovania slov sme porovnali so štandardnými slovnými modelmi a modelmi tried odvodených od slovných lém a slovných druhov. Priebežné experimenty ukazujú uspokojivé výsledky presnosti systému na rozpoznávanie reči v úlohe automatického prepisu spravodajských relácií.**

Kľúčové slová **– modelovanie jazyka, modely založené na triedach slov, rozpoznávanie plynulej reči, zhlukovanie slov**

I. Úvod

Jedným zo základných problémov pri štatistickom modelovaní vysoko-flektívnych alebo aglutinačných jazykov je *nedostatočnosť textových dát* (v angl. „*data sparsity*“), použitých pri trénovaní jazykových modelov. Obzvlášť je to v prípade slovenčiny, ktorá trpí nedostatkom rečových databáz a zdrojov textových dát. Okrem bohatej morfológie, ktorá generuje veľký počet ohybných tvarov slov, sa slovenčina vyznačuje tiež relatívne voľnou stavbou vetných štruktúr, čo spôsobuje zníženie prediktívnych schopností štandardných n-gramových modelov pri štatistickom modelovaní jazyka. Nárast počtu slov v slovníku vytvára tiež priestor na zvyšujúci sa počet slovných kontextov, ktoré sa v jazyku vyskytujú pomerne zriedkavo. To má za následok vzrastajúci počet n-gramov vyšších úrovní v jazykovom modeli, perplexita (zložitosť) modelu stúpa a odhad pravdepodobnosti takýchto n-gramov je zvyčajne nevierohodný. Preto je často nevyhnutné upriať pozornosť na sofistikovanejšie riešenia, ktoré eliminujú spomenuté nevýhody n-gramových modelov a zavádzajú do štatistického modelovania dodatočné syntaktické alebo sémantické závislosti medzi slovami, vychádzajúce priamo z morfológie daného jazyka. Jedným z možných riešení je *zhlukovať slová do tried* (v angl. „*word clustering*“) a modelovať jazyk pomocou slovných tried.

Modely založené na triedach slov (v angl. „*class-based models*“) sú pomerne široko uplatňované v rôznych aplikáciách zameraných na rozpoznávanie reči a spracovanie prirodzeného jazyka. Narozdiel od štandardných modelov, na triedach založené modely eliminujú problém nedostatočnosti trénovacích dát zhlukovaním slov do slovných tried, a tým redukujú celkový počet parametrov modelu [1] s cieľom efektívne vyhladiť a redukovať veľkosť jazykového modelu. Potom všetky slová umiestnené v jednej triede zdieľajú rovnaké štatistické závislosti vzhľadom k svojmu okoliu.

Existuje niekoľko rôznych spôsobov ako zhlukovať slová do tried. Prvotné techniky zvažovali *syntakticky-*, *sémanticky-* alebo *morfologicky-založené triedy* alebo *dátami-riadené prístupy zhlukovania slov do tried* [2]. Tieto triedy boli odvodené zvyčajne od slovných lém alebo morfologických značiek, ktoré boli priradené slovám pomocou morfologického analyzátora, navrhnutého alebo prispôsobeného pre daný jazyk.

Súčasné modelovanie jazyka pomocou slovných tried je zamerané na aplikáciu priamočiarejších metód zhlukovania slov do tried a na metódy, ktoré nevyužívajú komplexné algoritmy zjednoznačnenia (v angl. „*disambiguation*“) spôsobu priradenia morfologickej značky slovu. V článku [3] autori navrhli efektívny distribučne-založený a dátami-riadený algoritmus zhlukovania slov do tried pre systémy na automatický preklad (v angl. „*machine translation*“), ktorý je založený na dobre známom kritériu maximálnej vierohodnosti s viacerými úpravami. Modely založené na triedach slov boli tiež aplikované na zvýšenie účinnosti systému hľadania správnej odpovede na otázku užívateľa (v angl. „*question-answering*“) použitím techniky zhlukovania výrazov [4]. V oboch prípadoch boli modely tried interpolované so štandardnými slovnými modelmi s cieľom zvýšiť presnosť systému. V systémoch na rozpoznávanie plynulej reči s veľkým slovníkom (v angl. „*large vocabulary continuous speech recognition*“, skr. LVCSR), boli modely tried využité napríklad pri modelovaní aglutinačného estónskeho jazyka, odvodené od kmeňov a koncoviek slov pomocou neriadenej segmentácie slov do morfém založenej na princípe minimálnej opisnej dĺžky slova [5]. V prípade vysoko-flektívnej arabčiny, triedy slov boli odvodené od slovného kmeňa pomocou množiny ručne zadefinovaných morfologických pravidiel [6]. Podobný prístup, rozšírený o identifikáciu koncovky slova, bol použitý aj pri modelovani flektívneho perzského jazyka [7].

V tomto článku, sme sa zamerali na *morfologicky-založený prístup zhlukovania slov do tried* a na modelovanie slovenského jazyka pomocou tried odvodených do kmeňov a koncoviek slov použitím nami vytvorených pravidiel na identifikáciu slovného kmeňa a koncovky. Navrhnutý spôsob sme porovnali so štandardnými referenčnými modelmi a modelmi tried odvodených od slovných lém a morfologických značiek popisujúcich slovné druhy a gramatické kategórie slov. Pri tvorbe modelov založených na lémach a slovných druhoch sme využili dva morfologické analyzátory: český morfologický analyzátor Morče [8] a nami navrhnutý morfologický značkovač Dagger [9].

Tento článok je organizovaný nasledovne. Nasledujúca kapitola je vstupom do modelovania jazyka pomocou n-gramových modelov a modelov založených na triedach slov, opisuje spôsoby morfologického značkovania slov v slovenčine a navrhnutý prístup zhlukovania slov do tried pomocou slovného kmeňa a koncovky slova. V tretej kapitole predstavíme jednotné nastavenie systému na rozpoznávanie plynulej reči použité pri tvorbe experimentov. Ďalšia kapitola podáva prehľad o dosiahnutých experimentálnych výsledkoch navrhnutého prístupu zhlukovania slov do tried oproti konvenčným modelom. V záverečnej kapitole budú spomenuté hlavné prínosy článku s ďalším smerovaním v oblasti modelovania slovenského jazyka.

II. Modelovanie jazyka

A. n-gramové modely

Vo všeobecnosti, jazykový model v systéme LVCSR stanovuje apriórnu pravdepodobnosť $P(W)$ postupnosti slov $W = \{w_1 w_2 \dots w_n\}$, ktorá napomáha dekóderu nájsť tu najpravdepodobnejšiu postupnosť slov, ktorá korešponduje s vyslovenou akustickou informáciou od užívateľa systému. Súčasné modelovanie jazyka je založené na použití n-gramových modelov, ktoré zvažujú štatistické závislosti medzi slovami pomocou nasledujúceho vzťahu [2]:

$$P(W) = \prod_{i=1}^{n} P(w_i|h_i) = \prod_{i=1}^{n} P(w_i|w_{i-n+1} \dots w_{i-1}), \tag{1}$$

kde $P(w_i|h_i)$ je podmienená pravdepodobnosť slova w_i v závislosti od histórie $h_i = \{w_{i-n+1} \dots w_{i-1}\}$, ktorá je určená $(n-1)$ predchádzajúcimi slovami. Ako už bolo spomenuté v úvodnej časti, jedným zo základných problémov n-gramových modelov je to, že textové dáta použité pri trénovaní jazykového modelu majú veľký vnútorný rozptyl, a to znemožňuje spoľahlivo odhadnúť všetky možné podmienené pravdepodobnosti $P(w_i|h_i)$. Na elimináciu tohto problému boli preto navrhnuté modely založené na triedach slov.

B. Modely založené na triedach slov

Modely tried zhlukujú slová do ekvivalentných tried a modelujú jazyk ako súčin dvoch podmienených pravdepodobností nasledovne [1]:

$$P_{CL}(W) = \prod_{c \in C} P(w_i|c_i) P(c_i|c_{i-n+1} \dots c_{i-1}), \tag{2}$$

kde $P(w_i|c_i)$ je pravdepodobnosť slova w_i v danej triede c_i a $P(c_i|c_{i-n+1} \dots c_{i-1})$ je podmienená n-gramová pravdepodobnosť medzi triedami slov. Použitím tohto prístupu môžeme efektívne redukovať počet parametrov v jazykovom modeli (počet n-gramov, veľkosť modelu a pod.), pretože počet tried je vždy menší než celkový počet slov v slovníku a pravdepodobnosť slov, ktoré nesú podobné vlastnosti alebo zákonitosti v jazyku, je akumulovaná v rámci prislúchajúcej triedy.

C. Morfologické značkovanie v slovenčine

Ako už bolo spomenuté v úvodnej časti, k tvorbe jazykových modelov založených na triedach slov odvodených od ich lém a morfologických značiek sme použili dva morfologické analyzátory.

Prvý z nich, český morfologický analyzátor[1] Morče, využíva štatistický prístup založený na kombinácii *skrytého Markovovho modelu* (v angl. „*hidden Markov model*“, skr. HMM), ktorý transformuje postupnosť slov na postupnosť morfologických značiek, so *spriemerneným perceptronom* (v angl. „*averaged perceptron*“) slúžiaceho na odhad hodnôt prechodových váh medzi stavmi HMM [8]. Algoritmus využíva kombináciu 19 ručne vytvorených pravidiel zahrňujúcich medzislovný kontext a slúžiacich na výpočet parametrov perceptronu, resp. jeho spriemernenej hodnoty. Tento algoritmus bol neskôr prispôsobený pre potreby slovenčiny Slovenským národným korpusom (SNK) s využitím vlastnej množiny morfologických značiek (ďalej len značiek) [10].

Základným modulom druhého morfologického značkovača Dagger je klasifikátor založený na HMM, ktorý využíva súbor trigramových početností z ručne morfologicky anotovaného korpusu textových dát zo SNK [11]. HMM prvého rádu je opísaný maticou prechodov, ktorá je definovaná odhadom pravdepodobností značiek v závislosti od predošlej značky a maticou pozorovaní, ktorá je určená odhadom pravdepodobností značiek od aktuálneho slova. Tento základný model je doplnený o druhý HMM, ktorý spresňuje parametre prvého modelu. Doplňujúca matica pozorovaní je určená odhadom pravdepodobností značiek v závislosti od koncovky slova a matica prechodov obsahuje odhady pravdepodobností značiek od predchádzajúcej zjednodušenej značky. Výsledné odhady pravdepodobností sú napokon tvorené lineárnou kombináciou základného a doplňujúceho HMM, pričom doplňujúci model má nastavenú pomerne malú váhu, zistenú empirickým pozorovaním. V oboch prípadoch bol na určenie tej najpravdepodobnejšej postupnosti značiek využitý štandardný Viterbiho algoritmus. Vylepšením tohto princípu je využitie ručne morfologicky anotovaného slovníka z databázy SNK [11], ktorý obmedzuje prehľadávací priestor len na tie kombinácie „*slovo-značka*“, ktoré sa v danom slovníku vyskytujú [9].

D. Od morfém k triedam slov

Z jazykovedného hľadiska, flexia slovenského jazyka leží na hranici slovného kmeňa a koncovky. Pri tvorbe morfémových modelov slovenského jazyka sme na základe tohto poznatku vytvorili množinu pravidiel na *automatickú segmentáciu slov v tvare „kmeň-koncovka“* [12]. Rovnaký princíp sme neskôr použili aj pri zhlukovaní slov do tried na základe ich kmeňa alebo koncovky. Tento spôsob nám dovoľuje pomerne jednoducho a efektívne zhlukovať slová do tried bez potreby komplexného algoritmu na zjednoznačnenie, ako je to v prípade lematizácie alebo morfologického značkovania. Naviac daný spôsob zabezpečuje, že každé slovo v trénovacej množine bude zadelené len do jednej triedy (spĺňa požiadavky IBM zhlukovania [1]).

Navrhnuté pravidlá na automatickú identifikáciu a segmentáciu slovenských slov na kmene a koncovky je možné zhrnúť do nasledujúcich bodov:

- pravidlá na segmentáciu slov na kmene a koncovky vychádzajú z pravidiel slabičnej segmentácie v slovenčine [13], kde každá koncovka je reprezentovaná poslednou slabikou slova;
- segmentované sú len slová dĺžky viac než dve slabiky, resp. sedem písmen, zvyšné slová nepodliehajú pravidlám segmentácie;
- dĺžka koncovky je v rozsahu dvoch až štyroch písmen, resp. dĺžky jednej slabiky;
- segmentácia slov je obmedzená slovníkom, ktorý obsahuje približne 1 milión gramaticky správnych tvarov slov.

Ako bolo uvedené vyššie, takto navrhnuté pravidlá boli následne použité pri zhlukovaní slov do prislúchajúcich tried. Nesegmentované slová vystupovali v slovníku systému LVCSR ako slová, ktoré prináležia len jednej, vlastnej triede (slovo ako trieda).

III. Nastavenie systému na rozpoznávanie plynulej reči

Experimenty boli vykonávané s bigramovými a trigramovými modelmi, vytvorenými pomocou programovej sady SRILM [14], obmedzené slovníkmi veľkosti od $25k$ do $150k$ (s krokom $25k$) najviac frekventovaných slov, podmienených ich lemou, morfologickou značkou, slovným kmeňom alebo koncovkou slova (pozri Tabuľku 1 a Tabuľku 2). Na vyhladenie jazykových modelov bol použitý štandardný Wittenov-Bellov algoritmus v ústupovej schéme. Jazykové modely boli trénované na korpuse novinových článkov v rozsahu približne 180 miliónov tokenov, ktoré boli získané zo slovenských internetových stránok v rokoch 2007 až 2011 použitím nami vytvoreného systému na dolovanie, organizáciu a spracovanie textových dát nazývaného webAgent [15].

[1]Morfologickým analyzátorom sa označuje nástroj na lematizáciu a morfologické značkovanie slov obsahujúci pravidlá zjednoznačnenia pri výbere tej najpravdepodobnejšej značky pre dané slovo, odvíjajúce sa od okolitého kontextu.

Tabuľka 1
Porovnanie účinnosti modelov tried založených na lemách s modelmi tried odvodených od slovného kmeňa k referenčným celoslovným jazykovým modelom

ID	počet				OOV	bigramy		trigramy	
	slov	unigr.	bigr.	trigr.	%	PPL	WER %	PPL	WER %
referencia									
25k	25 046	25 046	10 448 110	11 926 759	13.42	495.88	31.80	395.90	30.69
50k	50 140	50 140	14 647 391	14 208 553	8.59	589.54	23.37	465.25	22.24
75k	75 230	75 230	16 964 041	15 220 704	6.73	637.80	20.28	501.00	19.31
100k	110 293	110 293	18 438 264	15 771 964	5.78	667.30	18.77	523.54	17.84
125k	125 601	125 601	19 466 441	16 110 118	5.19	691.04	17.90	542.16	16.95
150k	151 981	151 981	20 232 779	16 337 114	4.79	709.00	17.30	556.44	16.31
triedy odvodené od slovných lém - morfologický analyzátor Morče									
25k	25 038	8 762	6 423 854	12 317 554	9.29	387.72	34.20	298.65	35.14
50k	50 123	15 715	8 744 039	14 004 940	6.08	426.21	27.15	324.34	26.04
75k	75 099	22 262	9 946 892	14 672 965	4.95	447.20	24.35	339.71	23.20
100k	100 289	28 735	10 712 485	15 024 887	4.32	459.92	23.03	349.10	21.77
125k	125 873	34 915	11 218 327	15 225 721	3.88	470.02	22.05	356.71	20.85
150k	152 600	41 195	11 583 160	15 357 963	3.65	475.89	21.48	361.15	20.31
triedy slov odvodené od slovného kmeňa									
25k	25 027	16 265	10 116 595	13 241 250	9.20	494.79	31.06	382.86	29.81
50k	50 103	29 653	13 510 699	15 215 817	5.07	567.78	22.36	434.45	21.21
75k	75 322	42 702	15 329 118	16 069 378	3.55	604.53	19.56	461.62	18.51
100k	100 704	55 534	16 465 344	16 528 774	2.69	624.99	18.02	476.38	16.86
125k	126 320	68 011	17 222 162	16 805 118	2.20	641.53	17.11	488.45	16.01
150k	150 456	76 431	17 635 907	16 940 330	1.97	650.20	16.71	495.08	15.55

Použitý bol tiež trifónový kontextovo-závislý akustický model vytvorený na báze HMM, ktorého každý stav bol modelovaný zmesou 32 Gaussových funkcií. Model bol generovaný z príznakových vektorov, tvorených 39 mel-frekvenčnými kepstrálnymi koeficientami (v angl. „*mel-frequency cepstral coefficients*“, skr. MFCC). Natrénovaný bol na anotovaných rečových nahrávkach v rozsahu približne 60 hodín profesionálneho diktátu, získaného z databázy spravodajských relácií v slovenčine, zaznamenanej v rokoch 2007 až 2009. Táto akustická databáza je charakteristická vyváženosťou pohlavia rečníkov a obsahuje čítaný a spontánny prejav. Na modelovanie nepozorovaných trifón bol použitý efektívny algoritmus mapovania trifón, opísaný v [16].

Pri samotnom rozpoznávaní reči bol použitý LVCSR dekodér Julius, ktorý je založený na dvojprechodovej prehľadávacej stratégii, kedy vstupné dáta v prvom prechode sú spracované vždy pomocou bigramového jazykového modelu a výsledné hypotézy po rozpoznaní sú následne dotvorené reverzným n-gramovým modelom najvyššieho rádu [17].

Testovaciu množinu dát reprezentovali rečové nahrávky v rozsahu približne 240 minút získané náhodným výberom segmentov z už spomenutej databázy spravodajských relácií. Tieto segmenty reči neboli použité pri trénovaní akustického modelu a obsahujú 40 656 slov v 4 343 vetách.

Pri hodnotení jazykových modelov boli použité dve základné miery vychádzajúce z chybovosti systému LVCSR (v angl. „*word error rate*“, skr. WER) a zložitosti modelu (v angl. „*perplexity*“, skr. PPL), vyhodnocované na množine testovacích dát. WER je štandardná extrinzická miera na hodnotenie chybovosti systému LVCSR a udáva sa ako pomer hypotéz textu po rozpoznaní, ktorý v sebe zahŕňa chyby zmazaním, vložením a chyby zámenou, k referenčnému textu. Naopak pri intinzickom hodnotení jazykového modelu sa využíva miera zložistosti, tzv. perplexity modelu, ktorá udáva priemerný počet slov, medzi ktorými sa model v danom okamihu testovania rozhoduje a je definovaná ako reciproká hodnota váženého (geometrického) priemeru pravdepodobnosti určenú modelom pre každé slovo v testovacích dátach.

IV. Exprimentálne výsledky

Experimenty boli zamerané na hodnotenie kvality jazykových modelov pomocou perplexity modelu a chybovosti systému LVCSR vyhodnocovaných na množine testovacích dát, s cieľom preskúmať vplyv navrhnutej metodiky zhlukovania slov do tried v jazykovom modelovaní na celkovú úspešnosť systému LVCSR ako celku a pamäťové nároky aplikácie, v úlohe prepisu spravodajských relácií v slovenčine. Pri vyhodnocovaní kvality jazykových modelov sme sa zamerali na tri druhy experimentov. Vo všetkých prípadoch sme modely založené na triedach slov porovnávali s referenčnými celoslovnými jazykovými modelmi.

V prvom prípade sme zisťovali vplyv modelovania slovenského jazyka pomocou tried odvodených od slovných kmeňov slov oproti konvenčným modelom založených na slovných lemách. Z jazykovedného hľadiska je prípad lematizácie a hľadania slovného kmeňa (v angl. „*stemming*“) veľmi podobný, pretože kmeňové morfémy odrážajú lexikálny význam slova, ktoré po rozšírení o koncovku v základnom tvare vytvárajú slovné lemy. Ako je možné vidieť v Tabuľke 1, modely založené na lemách vykazujú mierne zhoršené výsledky chybovosti systému LVCSR, blížiace sa

Tabuľka 2
Porovnanie účinnosti modelov tried založených na slovných druhoch s modelmi tried odvodených od koncovky slova k referenčným celoslovným jazykovým modelom

ID	počet				OOV	bigramy		trigramy	
	slov	unigr.	bigr.	trigr.	%	PPL	WER %	PPL	WER %
referencia									
25k	25 046	25 046	10 448 110	11 926 759	13.42	495.88	31.80	395.90	30.69
50k	50 140	50 140	14 647 391	14 208 553	8.59	589.54	23.37	465.25	22.24
75k	75 230	75 230	16 964 041	15 220 704	6.73	637.80	20.28	501.00	19.31
100k	110 293	110 293	18 438 264	15 771 964	5.78	667.30	18.77	523.54	17.84
125k	125 601	125 601	19 466 441	16 110 118	5.19	691.04	17.90	542.16	16.95
150k	151 981	151 981	20 232 779	16 337 114	4.79	709.00	17.30	556.44	16.31
triedy odvodené od slovných druhov - morfologický analyzátor Morče									
25k	25 030	698	197 798	2 044 236	0.46	76.68	41.47	70.61	41.39
50k	50 110	820	225 673	2 142 625	0.20	75.62	31.57	69.48	31.00
75k	75 023	881	235 064	2 165 494	0.16	74.84	27.92	68.74	27.22
100k	100 003	927	240 552	2 177 163	0.13	74.32	25.90	68.23	25.33
125k	125 833	961	244 307	2 183 768	0.13	73.91	24.61	67.86	23.88
150k	151 865	1 004	247 632	2 189 237	0.12	73.68	23.60	67.65	22.88
triedy odvodené od slovných druhov - morfologický značkovač Dagger									
25k	25 016	842	200 243	2 298 505	0.45	85.47	37.86	79.93	37.33
50k	50 035	970	224 660	2 381 232	0.26	87.80	28.23	81.99	27.77
75k	75 371	1 033	233 027	2 402 723	0.20	87.89	24.32	82.04	23.75
100k	100 564	1 085	239 061	2 414 019	0.17	87.49	22.30	81.66	21.44
125k	125 923	1 118	242 194	2 419 433	0.16	87.29	21.39	81.48	20.48
150k	151 540	1 148	244 927	2 424 014	0.14	87.05	20.45	81.26	19.74
triedy slov odvodené od koncovky slova									
25k	25 000	9 830	4 916 168	12 797 644	3.52	300.98	30.77	240.20	29.52
50k	50 133	16 538	6 034 804	13 579 290	2.19	315.52	22.17	250.91	20.97
75k	75 336	22 794	6 615 068	13 887 128	1.72	321.29	19.24	255.07	18.13
100k	100 720	28 945	6 985 999	14 055 042	1.44	324.90	17.62	257.74	16.43
125k	126 337	35 088	7 246 875	14 159 824	1.28	327.56	16.83	259.61	15.59
150k	150 474	38 075	7 347 131	14 194 938	1.23	328.48	16.24	260.38	14.95

k hodnote 20% WER. Naopak, hodnoty perplexity a percenta mimoslovníkových tvarov slov (v angl. „*out-of-vocabulary*“, skr. OOV) a veľkosti modelu výrazne klesli, priemerne o tretinu, najmä z dôvodu výraznej redukcie počtu tried (unigramov) v jazykovom modeli. Čo sa týka modelov tried odvodených od slovných kmeňov, dosiahli sme určitý kompromis v hodnotách perplexity, v počte tried a OOV slov v slovníku vzhľadom k celoslovným modelom a modelom založeným na lemách. V tomto prípade sme však zaznamenali aj mierny pokles chybovosti systému LVCSR, relatívne o 3.5% WER v prípade bigramov a 4.5% WER v prípade trigramov. Z výsledkov možno usúdiť, že modely založené na triedach slov odvodených od slovných kmeňov prinášajú mierne zlepšenie vo všetkých hodnotiacich mierach oproti konvenčným prístupom.

Druhá časť experimentov bola zameraná na zistenie vplyvu modelovania slovenského jazyka pomocou tried odvodených od koncoviek slov oproti modelom založených na slovných druhoch, reprezentovaných morfologickými značkami. Ako je možné vidieť v Tabuľke 2, v oboch prípadoch modelov založených na slovných druhoch možno pozorovať výrazné zníženie perplexity a hodnoty percenta OOV slov. Táto skutočnosť je však daná výpočtom hodnôt perplexity a percenta OOV, ktorá bola vyhodnocovaná iba medzi triedami slov, nie medzi jednotlivými slovami. Hodnoty chybovosti systému LVCSR však výrazne vzrástli, v priemere o pätinu hodnoty WER v prípade značkovania slov pomocou morfologického značkovača Dagger a o tretinu v prípade priradzovania značiek pomocou morfologického analyzátora Morče. Hlavným prínosom daného prístupu zhlukovania slov do tried je však výrazná redukcia počtu n-gramov v jazykovom modeli, a tým aj výslednej veľkosti jazykového modelu, pohybujúcej sa v rozmedzí od 6 do 17 MB v prípade bigramov a od 60 do 90 MB pri trigramoch (oproti referenčným modelom o veľkosti v rozmedzí od 270 do 540 MB v prípade bigramov a od 650 do $1\,100$ MB pri trigramoch), čo záviselo najmä od počtu tried obsiahnutých v jazykovom modeli a malo výrazný vplyv na zníženie pamäťových nárokov výslednej aplikácie v rozpoznávaní reči. Porovnateľným prístupom zhlukovania slov do tried k slovným druhom je zhlukovať slová na základe ich koncovky, pretože ako už bolo spomenuté v úvode článku, veľká časť flexie slovenského jazyka je sústredená práve na konci slova a táto v prevažnej väčšine odráža gramatickú kategóriu slova. Pri tomto prístupe došlo k zníženiu perplexity modelov približne o tretinu v porovnaní s referenčnými modelmi a percentuálna hodnota OOV klesla o viac než 75%. Ak sa pozrieme na hodnoty chybovosti systému LVCSR, zaznamenali sme ďalšie mierne zníženie hodnoty WER približne o 5% v prípade bigramov a o viac než 6% v prípade trigramov, v relatívnej miere. Na základe vykonaných experimentov možno konštatovať, že modely založené na triedach slov odvodené od koncoviek slov prinášajú mierne zlepšenie presnosti systému LVCSR oproti všetkým predchádzajúcim prístupom a sú ako v predchádzajúcom

prípade určitým kompromisom medzi modelmi založenými na slovných druhoch a referenčnými modelmi, čo do počtu tried, percenta OOV, či hodnoty perplexity modelov. Z týchto dôvodov je vhodné nasledujúci výskum v oblasti modelovania slovenského jazyka zamerať na morfologické modely odvodené od kmeňov a koncoviek slov, prípadne na ich vhodnú kombináciu.

Tretia časť experimentov bola zameraná na vyhodnotenie účinnosti priradzovania značiek pomocou morfologického analyzátora Morče a značkovača Dagger. Na základe vykonaných experimentov možno povedať, že aj napriek mierne zvýšenému počtu tried v jazykovom modeli, ktorý sa odrazil aj vo výslednej hodnote perplexity modelov, nami navrhnutý spôsob značkovania pomocou nástroja Dagger bol priemerne o 30% v prípade bigramov a o 12% v prípade trigramov účinnejší než analyzátor Morče, aj napriek tomu, že morfologicky značené dáta použité pri trénovaní značkovača Dagger vychádzali z dát označkovaných analyzátorom Morče.

V. Záver

V tomto článku boli predstavené a experimentálne overené morfologicky-založené modely slovenského jazyka využívajúce triedy slov vytvorené zhlukovaním slov na základe ich kmeňa a koncovky. Kategorizácia slov do tried vychádzala z morfologických pravidiel navrhnutých pre automatickú segmentáciu slov na kmene a koncovky v slovenčine. Na základe experimentálnych výsledkov je možné usúdiť, že morfologické triedy odvodené od kmeňov a koncoviek slov môžu priniesť potenciálne zlepšenie kvality jazykového modelu, vzhľadom na perplexitu modelu, veľkosť modelu a počet mimoslovníkových tvarov slov, ako aj samotné zvýšenie presnosti systému na automatické rozpoznávanie plynulej reči ako celku, oproti konvenčným metódam zhlukovania slov na základe slovných lém alebo slovných druhov a modelovaním jazyka pomocou týchto tried. Tento poznatok bude tiež predmetom ďalšieho výskumu zameraného na aplikáciu navrhnutého spôsobu zhlukovania slov do tried v reálnom doménovo-orientovanom rozpoznávaní plynulej reči so zreteľom na slovenský jazyk.

Poďakovanie

Táto práca bola vytvorená realizáciou projektu Centrum informačných a komunikačných technológií pre znalostné systémy (kód ITMS projektu: 26220120020) na základe podpory operačného programu Výskum a vývoj, financovaného z Európskeho fondu regionálneho rozvoja (100%).

Literatúra

[1] P. F. Brown, V. J. Della Pietra, P. V. DeSouza, J. C. Lai and R. L. Mercer, *Class-Based n-gram Models of Natural Language*, In: Computational Linguistics, Vol. 18, No. 4, pp. 467–479, ISSN 0891-2017, 1992.

[2] D. Jurafsky and J. H. Martin, *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition* (2^{nd} *edition*), Prentice Hall, Pearson Education, pp. 998, ISBN 978-0-13-504196-3, 2009.

[3] J. Uszkoreit and T. Brants, *Distributed Word Clustering for Large Scale Class-Based Language Modeling in Machine Translation*, In: Proc. of the 46^{th} Annual Meeting of the ACL-08: HLT, Columbus, Ohio, USA, pp. 762–775, 2008.

[4] S. Momtazi and D. Klakow, *A Word Clustering Approach for Language Model-Based Sentence Retrieval in Question Answering Systems*, In: Proc. of the 18^{th} ACM Conference on Information and Knowledge Management, CIKM'2009, Hong Kong, China, pp. 1911–1914, 2009.

[5] T. Alumae, *Large Vocabulary Continuous Speech Recognition for Estonian using Morpheme Classes*, In: Proc. of ICSLP'2004, Jeju Island, Korea, pp. 389–392, 2005.

[6] A. Ghaoui, F. Yvon, Ch. Mokbel and G. Chollet, *On the Use of Morphological Constraints in n-gram Statistical Language Modeling*, In: Proc. of INTERSPEECH'2005, Lisbon, Portugal, pp. 1281–1284, 2005.

[7] H. Faili and H. Ravanbakhsh, *Affix-Augmented Stem-Based Language Model for Persian*, In: Proc. of the 6^{th} IEEE International Conference on NLP-KE'2010, Beijing, China, pp. 250–253, 2010.

[8] D. Spoustová, J. Hajič, P. Krbec, P. Květoň and J. Votrubec, *The Best of Two Worlds: Cooperation of Statistical and Rule-Based Taggers for Czech*, In: Proc. of Balto-Slavonic Natural Language Processing Workshop, Prague, Czech Republic, pp. 67–74, 2007.

[9] D. Hládek, J. Staš and J. Juhár, *Dagger: The Slovak Morphological Classifier*, In: Proc. of the 54^{th} International Symposium ELMAR-2012, Zadar, Croatia, pp. 1–4, (to be published), 2012.

[10] M. Šimková, *Slovak National Corpus - History and Current Situation*, In: M. Šimková (Ed.): Insight into Slovak and Czech Corpus Linguistics, Veda, Bratislava, Slovak Republic, pp. 166–175, 2005.

[11] JUĽŠ-SNK, *Ručne morfologicky anotovaný korpus*, Jazykovedný ústav Ľudovíta Štúra, Slovenský národný korpus, Bratislava, Slovenská republika, 2007. URL: http://korpus.juls.savba.sk/stats.html

[12] D. Zlacký, *Morfémový model slovenského jazyka v systémoch automatického rozpoznávania reči*, Diplomová práca, Katedra elektroniky a multimediálnych telekomunikácií, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika, s. 54, 2012.

[13] J. Ivanecký, *Štatistický prístup pri určovaní slabičných hraníc*, In: Proc. of SLOVKO'2003, Bratislava, Slovak Republic, pp. 11, (in Slovak), 2003.

[14] A. Stolcke, *SRILM – An Extensible Language Modeling Toolkit*, In: Proc. of ICSLP'2002, Denver, Colorado, USA, pp. 901–904, 2002.

[15] D. Hládek and J. Staš, *Text Mining and Processing for Corpora Creation in Slovak Language*, In: Journal of Computer Science and Control Systems, Vol. 3, No. 1, pp. 65–68, ISSN 1844-6043, 2010.

[16] S. Darjaa, M. Ceňak, M. Trnka and M. Rusko, *Effective Triphone Mapping for Acoustic Modeling in Speech Recognition*, In: Proc. of INTERSPEECH'2011, Florence, Italy, pp. 1717–1720, 2011.

[17] T. Lee, T. Kawahara and K. Shikano, *Julius – An Open Source Real-Time Large Vocabulary Recognition Engine*, In: Proc. of EUROSPEECH'2001, Aalborg, Denmark, pp. 1691–1694, 2001.

Evaluácia difónovej syntézy reči z textu slovenského jazyka

Martin SULÍR, Jozef JUHÁR

Katedra elektroniky a multimediálnych telekomunikácií, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

martin.sulir@student.tuke.sk, jozef.juhar@tuke.sk

***Abstrakt* — Táto práca sa zaoberá problematikou tvorby nových slovenských hlasov (mužského a ženského) pre difónovú syntézu reči z textu za použitia systému prevodu textu na reč – Festivalu a ich následnou evaluáciou pomocou objektívnych a subjektívnych metód hodnotenia kvality syntetických hlasov. Festival predstavuje kompletný TTS systém podporujúci ako spracovanie a analýzu vstupného textu, tak aj finálnu syntézu rečového signálu na výstupe. TTS systémy, ktoré sú v tejto práci navrhnuté, pozostávajú z vytvorenia novej akustickej databázy pre Festival a jej následným spracovaním za pomoci príslušných podprogramov a skriptov, čo zabezpečuje integráciu slovenského jazyka priamo do systému. Pre evaluáciu novovytvorených hlasov boli použité subjektívne miery hodnotenia MOS a MRT spoločne s objektívnou metódou PESQ.**

***Kľúčové slová* — Syntéza reči, TTS systém, Festival, Difónová syntéza**

I. Úvod

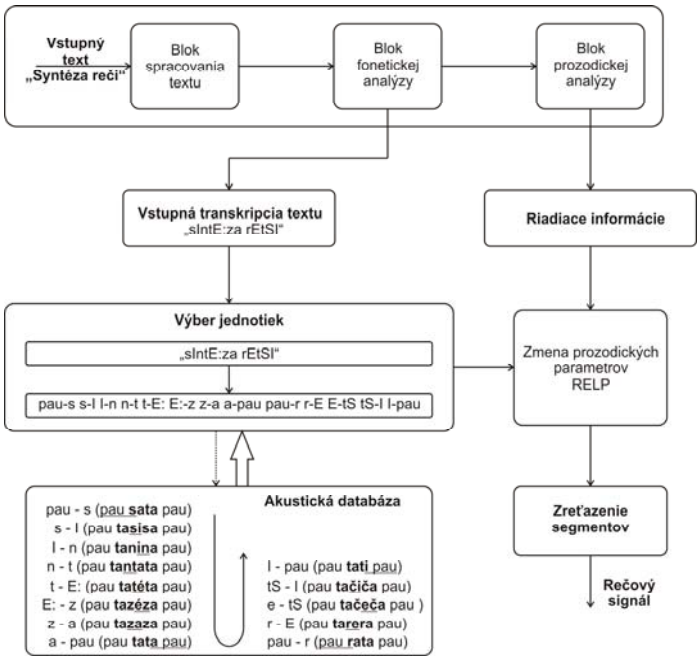
Syntézu reči v dnešnej dobe, v ktorej je výpočtová technika braná ako bežná súčasť života, predstavujú najmä počítačové systémy, ktoré dôkazu prevádzať vstupný text na reč. Základnou úlohou takýchto systémov je v prvom rade uľahčenie života, či už ľuďom s telesným hendikepom ako napríklad nevidiacim alebo aj úplne bežným ľuďom, ktorým využitie týchto systémov uľahčuje dennodenné úkony. Výskum v tejto oblasti smeruje k bodu, kedy bude možné takéto „umelé“ hlasy využívať masovo v rôznych oblastiach bez toho, aby boli akosi limitované a pôsobili umelo, ba až neprirodzene.

Jednotlivé metódy syntézy reči môžeme rozdeliť podľa viacerých kritérií, akými môžu byť napríklad rozdelenie podľa oblasti, v ktorej syntéza reči prebieha (frekvenčná alebo časová oblasť), podľa spôsobu riadenia syntézy (riadená pravidlami alebo dátami), alebo rozdelenie podľa prístupu k tvorbe rečového signálu. Rozdelenie metód syntézy reči podľa prístupu k tvorbe rečového signálu pozostáva zo syntézy reči založenej na modelovaní vokálneho traktu, konkatenačnej syntéze a štatistickej parametrickej syntéze [1]. Tento článok je zameraný na problematiku tvorby a následnej evaluácie nových TTS systémov s využitím konkatenačnej metódy syntézy reči založenej na spájaní fonetických elementov – difón. Tvorbu týchto systémov je možné rozdeliť do dvoch základných krokov, ktorými sú tvorba akustickej databázy a jej následná implementácia pomocou softvérového nástroja Festival [2].

II. Difónová syntéza reči z textu

Metóda difónovej syntézy reči využíva najpoužívanejšie fonetické elementy pre syntézu - difóny [3]. Difóny predstavujú prípustný kompromis medzi ich početnosťou a možnou dosiahnutou kvalitou syntézy. Základnou vlastnosťou difón, ktorá ich predurčuje pre použitie v systémoch spracovania reči je ich konštrukcia, keďže sú tvorené oblasťou od stredu prvej hlásky do stredu nasledujúcej hlásky, a teda ako začiatok, tak aj ich koniec sa nachádza v najstabilnejšej oblasti a nie v oblasti prechodu medzi hláskami, ktorá obsahuje najviac užitočných informácií. Maximálny počet difón daného jazyka je odvodený od počtu foném, takže ak uvažujeme pre slovenčinu 53 foném výsledný počet difón bude $53*53=2809$. Avšak ak vylúčime kombinácie foném, ktoré sa v slovenčine nevyskytujú výsledný počet difón bude v rozmedzí okolo 1200 až 1600, čo predstavuje pomerne jednoducho zrealizovateľný základ pre databázu obsahujúcu potrebné difóny daného jazyka.

Princíp difónovej syntézy (Obr.1) spočíva v prvom kroku z výberu požadovaných difón z akustickej databázy slovenského jazyka. Táto databáza obsahuje nahrávky nosných slov všetkých potrebných difón a dodatočné informácie, akými sú napríklad informácia o umiestnení jednotlivých difón v nahrávkach alebo informácia o značkách periód (pitchmark). Ďalším krokom je modifikácia prozodických parametrov. Pre túto modifikáciu je možné využiť viacero algoritmov, ako napríklad MBROLA, RELP alebo TD PSOLA. Vo Festivale je implementovaná práve metóda RELP, ktorá využíva pre modifikáciu prozodických parametrov LPC koeficienty a ich rezídua získané z nahrávok. Segmenty s upravenými parametrami sú následne zreťazené, čím sa získa syntetická reč.



Obr. 1 Bloková schéma difónovej syntézy reči

III. METÓDY PRE VYHODNOCOVANIE KVALITY SYNTÉZY REČI

Pre evaluáciu TTS systémov je možné vo všeobecnosti použiť dva typy hodnotení [4]. Prvou skupinou sú subjektívne hodnotenia, ktoré spočívajú v posudzovaní kvality syntézy reči používateľmi, ktorým sú predkladané ukážky syntézy a ich úlohou je na základe predkladaných dotazov ohodnotiť tieto ukážky. Druhou skupinou sú objektívne hodnotenia, ktoré spočívajú v aplikácií metód vyhodnocovania kvality založených na psychoakustickom modelovaní sluchových ústrojov človeka a na kognitívnom modelovaní rozhodovania o počuteľných vnemoch vykonávaného ľudským mozgom.

A. Subjektívne metódy hodnotenia kvality reči

V tejto práci boli využite dva typy subjektívnych hodnotení a to MOS a MRT. Testovanie MOS (Mean opinion score) predstavuje subjektívnu mieru pre celkové testovanie kvality reči [5]. Tento typ testu spočíva v ohodnotení jednotlivých parametrov reči pomocou mierky od 1 do 5 a následnom vypočítaní aritmetického priemeru získaných hodnotení. Mierka pre hodnotenie kvality jednotlivých parametrov teda pozostáva z piatich hodnôt, ktorým je priradený stupeň kvality a to tak, že hodnota 1 odpovedá zlej kvalite a hodnota 5 predstavuje vynikajúcu kvalitu hodnoteného parametra. MRT test patrí medzi jeden z testov, ktorý slúži pre hodnotenie zrozumiteľnosti reči [5]. Test pozostáva z identifikácie a následnej voľby správnej možnosti zo šiestich ponúknutých setov slov, pričom každý z týchto setov obsahuje šesť jednoslabičných foneticky podobných slov líšiacich sa v prvej alebo poslednej spoluhláske. Pri testovaní sa užívateľovi prehrá jeden zo setov s určitým poradím slov a on musí následne vybrať jednu z možností (uzavretá odpoveď), v ktorých sú tieto slová náhodné poprehadzované alebo musí dopísať, aké slová počul (otvorená odpoveď). Vyhodnocovanie testovania je možné pomocou percentuálnej úspešnosti výberu správneho setu, pričom prirodzený ľudský hlas dáva obvykle 99% úspešnosť správneho rozpoznania setu a pri syntéze reči sa tieto hodnoty pohybujú v rozmedzí od 70 do 95%.

B. Objektívne metódy hodnotenia kvality reči

Pre objektívne vyhodnotenie kvality reči bola využitá metóda PESQ (Perceptual Evaluation of Speech Quality). Metóda PESQ, ktorá bola popísaná v odporúčaní ITU-T P.862 predstavuje intruzívnu objektívnu metódu pre hodnotenie kvality reči na základe porovnávania referenčného a znehodnoteného signálu [6]. Princíp činnosti algoritmu PESQ spočíva zo spracovania oboch signálov, ich časového zarovnania a následnej aplikácií psychoakustického modelu, ktorý je zodpovedný za výpočet rozdielov medzi referenčným a znehodnoteným signálom. Výstupom tohto porovnávania je číslo z rozsahu od -0,5 do 4,5, kde -0,5 zodpovedá najnižšiemu stupňu kvality a 4,5 predstavuje maximálnu možnú kvalitu testovaného signálu. Keďže stupnica PESQ priamo nezodpovedá stupnici subjektívneho hodnotenia MOS a v mnohých prípadoch je potrebné tieto dva hodnotenia porovnávať, bola v odporúčaní ITU-T P.862.1 definovaná mapovacia funkcia, ktorá prispôsobuje získané hodnotenie z algoritmu PESQ stupnici MOS [7]. Mapovacia funkcia pre prevod PESQ hodnotenia do stupnice MOS má nasledujúci tvar:

$$y = 0.999 + \frac{4.999 - 0.999}{1 + e^{-1.4945 * x + 4.6607}}, \tag{1}$$

kde zodpovedá PESQ hodnoteniu a je tzv. MOS-LQO (Mean opinion score – Listening Quality Objective) hodnotenie, ktoré predstavuje PESQ po mapovaní do mierky MOS. Hodnotenie po mapovaní nadobúda čísla z rozsahu od 1 do 4,5.

V prípade hodnotenia kvality syntézy reči algoritmom PESQ predstavuje vzorka syntetizovanej reči znehodnotený signál a originálna nahrávka jeho referenciu. Nahrávky musia byť pravdaže obsahovo rovnaké a originálna nahrávka by mala byť nahratá totožným rečníkom ako východisková databáza testovaného TTS systému (v prípade TTS systémov využívajúcich akustickú databázu).

IV. POPIS EXPERIMENTOV

Evaluácia novovytvorených hlasov sa uskutočnila ako subjektívnymi, tak aj objektívnymi metódami. Subjektívne hodnotenie, čiže posluchové testy hlasov, spočívali vo vypočutí nahrávok a následnom vyplňovaní dotazníka užívateľmi. Tento dotazník bol vytvorený ako HTML stránka, ktorá zastrešovala oba kroky tohto testovania. HTML stránka sa skladala zo štyroch modulov pre prehrávanie nahrávok a štyroch dotazníkov, pričom každý hlas bol testovaný metódou MOS a metódou MRT. Prvou časťou evaluácie oboch hlasov bolo teda hodnotenie základných parametrov reči (zrozumiteľnosti, prirodzenosti, rýchlosti, plynulosti reči a príjemnosti hlasu) pomocou mierky MOS. Posudzovaná syntetizovaná nahrávka ako pre mužský, tak aj pre ženský hlas obsahovala nasledujúcu vetu: „Vážené dámy, vážení páni zaujmite prosím svoje miesta a vypnite mobilné telefóny. Premiéra muzikálu Rybár a jeho duša ochvíľu začína.“. Táto veta bola zvolená najmä preto, že bola použitá už ako ukážka hlasu (v prirodzenej podobe) vtedy ešte potencionálneho rečníka, ktorý následne nahrával akustickú databázu pre ženský hlas, a teda jej použitie bolo efektívne z hľadiska následného objektívneho testovania, kde poslúžila ako referenčný vstup do systému.

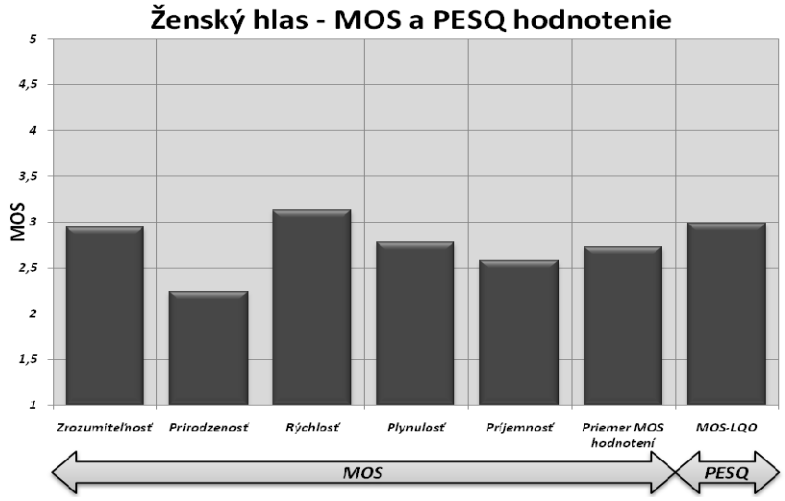
Druhou časťou testovania bol test založený na metóde MRT, kde sa testovala zrozumiteľnosť hlasov na základe vypočutia si nahrávky sady foneticky podobných slov a následnej voľby jednej z možností, ktorá obsahovala dané slová v správnom poradí. V tomto prípade sa nahrávky pre jednotlivé hlasy líšili, mužský hlas obsahoval nahrávku nasledujúcich slov: „bez, les, lez, des, ves, cez“ a ženský hlas nahrávku týchto slov: „zať, žať, hať, sať, mať, dať“. Odpovede v dotazníku tvorili dané slová v pozmenenom poradí.

Pre objektívne testovanie bola zvolená metóda PESQ. Toto testovanie sa uskutočnilo pomocou systému, ktorý je súčasťou oficiálnej distribúcie odporúčania ITU-T P.861. Do tohto systému vstupovali ako znehodnotené signály nahrávky syntetických hlasov a ako referencie nahrávky s totožným obsahom v prirodzenej podobe. Nahrávky museli byť prevzorkované z pôvodnej frekvencie 44,1 kHz na frekvenciu 16 kHz, keďže metóda PESQ podporuje iba vzorkovacie frekvencie nahrávok 8 a 16 kHz. Systém následne vyhodnotil kvalitu reči pomocou čísla z rozsahu daného pre túto metódu. Pre porovnanie získaných výsledkom musela byť táto hodnota premapovaná na hodnotenie MOS-LQO pomocou vzorca (1), ktoré umožňuje porovnávanie PESQ výsledkov s výsledkami subjektívnych MOS hodnotení.

V. VÝSLEDKY EXPERIMENTOV

A. Výsledky hodnotenia kvality ženského hlasu

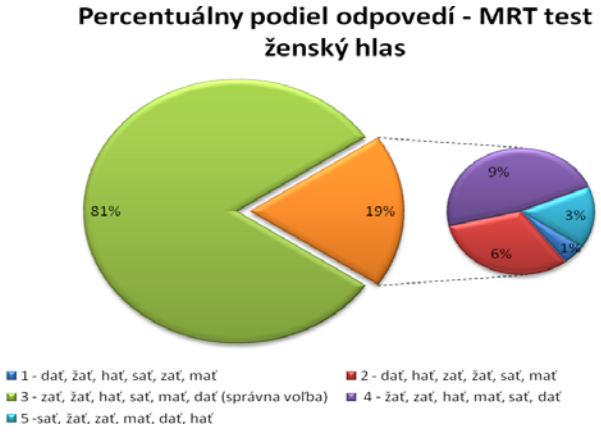
Subjektívneho hodnotenia kvality ženského hlasu pomocou mierky MOS sa zúčastnilo 101 používateľov, ktorý hodnotili jednotlivé parametre tohto hlasu. Nadobudnuté hodnoty pre dané parametre boli následne spriemerované pre získanie celkového hodnotenia systému, ktorý bolo možné porovnávať s objektívnym hodnotením. Objektívne hodnotenie bolo pre potreby porovnávania prepočítané na hodnoty MOS-LQO. Nadobudnuté výsledky sú zobrazené na Obr.2.



Obr. 2 Získané hodnotenia kvality reči MOS a PESQ ženského hlasu

Ako je z grafu zrejmé, najnižšie hodnotenie získala prirodzenosť reči. Toto nízke hodnotenie je spôsobené najmä využitím LPC v RELP metóde, ktorá do značnej miery limituje získanie požadovanej prirodzenosť reči. S nízkou prirodzenosťou reči je spojené aj nižšie hodnotenie príjemnosti hlasu, keďže práve prirodzenosť do značnej miery definuje príjemnosť hlasu. Ostatné parametre, ako aj ich celkový priemer boli v rozsahu od 2,5 do 3,5 čo predstavuje typický rozsah pre hodnotenie kvality syntetickej reči [5]. Ako je ďalej zrejmé, rozdiel medzi objektívnym hodnotením MOS-LQO a priemerom MOS hodnotení je minimálny, čím sa potvrdilo možné využitie objektívneho hodnotenia PESQ pre vyhodnocovanie kvality syntézy, aj keď subjektívne hodnotenie stále predstavuje dôležitú súčasť takýchto hodnotení, keďže spätná väzba od používateľov, pre ktorých je takýto TTS systém určený je veľmi dôležitá.

Druhým typom testu ženského hlasu bol test zrozumiteľnosti na spôsob metódy MRT. Tohto testu sa zúčastnilo taktiež 101 používateľov. Percentuálny podiel jednotlivých odpovedí je zobrazený na Obr.3.



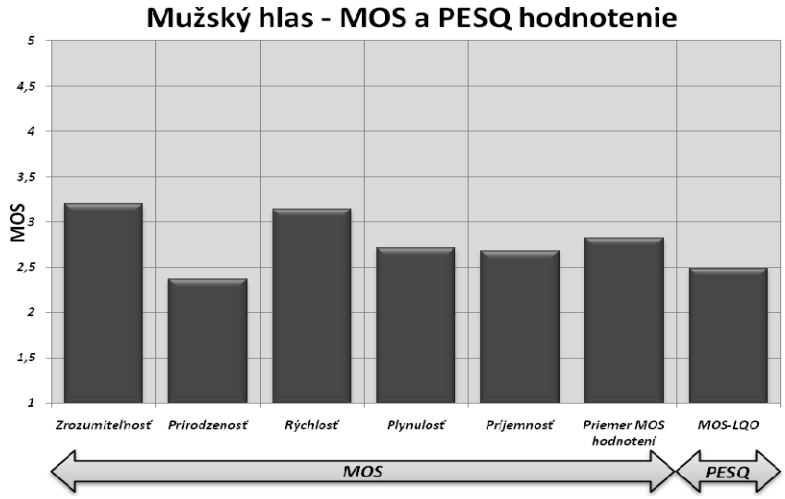
Obr. 3 Percentuálny podiel odpovedí MRT testu ženského hlasu

V tomto grafe predstavuje správna voľba setu slov 81% zo všetkých odpovedí. Najväčšiu časť zo zvyšných 19% tvorí odpoveď, v ktorej nastala zámena dvoch foneticky veľmi

podobných slov a je teda možné povedať, že aj keď bola zrozumiteľnosť tohto setu slov vyhodnotená ako veľmi dobrá, keďže obvyklé hodnoty sú pre tento test zrozumiteľnosti syntetickej reči z rozsahu od 70% do 95% [5], stále pretrvával problém porozumieť foneticky veľmi podobným slovám. Avšak pre získanie presnejšieho hodnotenia zrozumiteľnosti systému ako celku, by bolo nutné takýmto spôsobom otestovať oveľa väčšiu sadu setov. Tento test bol preto realizovaný ako experiment, pričom sa potvrdilo, že zrozumiteľnosť hlasu sa dá považovať za dobrú.

B. Výsledky hodnotenia kvality mužského hlasu

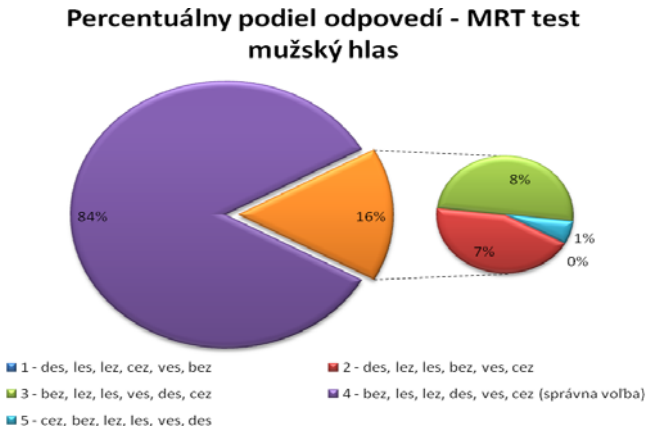
Rovnaké hodnotenie, ako to bolo v prípade ženského hlasu, bolo realizované aj pre mužský hlas. Tohto hodnotenia sa zúčastnilo 98 používateľov. Rozdielny počet používateľov bol spôsobený tým, že nie všetci vyplnili každý z predkladaných testov. Získané výsledky hodnotení mužského hlasu sú zobrazené na Obr.4.



Obr. 4 Získané hodnotenia kvality reči MOS a PESQ mužského hlasu

Ako je z grafu zrejmé, prirodzenosť tohto hlasu opäť získala najnižšie hodnotenie. Príčiny tejto nízkej hodnoty sú rovnaké, ako tomu bolo v prípade ženského hlasu, a teda využitie LPC v metóde RELP. Avšak v tomto prípade nízka prirodzenosť hlasu mala menší dopad na hodnotenie príjemnosti hlasu. Nižšie hodnotenie príjemnosti v prípade ženského hlasu mohlo byť teda spôsobené všeobecným spoločenským očakávaním, že ženský hlas bude pôsobiť príjemne a melodicky. Ostatné výsledky hodnotených parametrov, ako aj ich priemer sa opäť nachádzajú v rozsahu obvyklých hodnôt pre syntézu reči. Objektívne hodnotenie v tomto prípade vykazuje vyššiu odchýlku od priemeru MOS hodnotení, avšak stále potvrdzuje jej možné využitie ako jedného z kritérií pre hodnotenie TTS systémov.

Aj v prípade mužského hlasu obsahoval druhý dotazník test zrozumiteľnosti na spôsob MRT testu. Percentuálny podiel získaných odpovedí je zobrazený na Obr.5.



Obr. 5 Percentuálny podiel odpovedí MRT testu mužského hlasu

Z grafu je zrejmé, že správna voľba setu v tomto prípade znamenala 84% zo všetkých odpovedí. Opäť sa potvrdil jav, ktorý bol spozorovaný už pri hodnotení zrozumiteľnosti ženského hlasu, že dve najpočetnejšie odpovede z ostatných 16% predstavovali možnosti, v ktorých dochádzalo k zámene foneticky veľmi podobných slov. Z testu je tak isto možno dedukovať udržiavanie pozornosti pri počúvaní nahrávok, kde sa prejavilo, že najväčšia pozornosť sa udržiavala na začiatku nahrávok.

VI. ZÁVER

V tejto práci je prezentované využitie a následná evaluácia dvoch novovytvorených hlasov pre difónovú syntézu reči s použitím softvérového nástroja Festival pre slovenský jazyk. Subjektívnym testovaním sa ukázalo, že používatelia dané hlasy ohodnotili adekvátnym hodnotením na základe použitej metódy syntézy reči, čo dokázalo aj subjektívne hodnotenie oboch hlasov. Hlasy disponujú podľa získaných výsledkov prijateľnou zrozumiteľnosťou a plynulosťou reči na výstupe. Za týmito hodnoteniami značne zaostáva hodnotenie prirodzenosti reči, čo je spôsobené najmä využitím RELP metódy pri syntéze. Z hľadiska porovnania získaným hodnotení subjektívnych a objektívnych testovaní sa dá povedať, že využitie objektívnych metód pre evaluáciu TTS systémov je možné, avšak získané hodnotenia zdôrazňujú dôležitosť a opodstatnenosť subjektívnych testovaní, keďže TTS systémy sú vytvárané pre používateľov, a tak by mala do procesu tvorby takýchto systémov vstupovať spätná väzba práve od nich. Dá sa teda povedať, že na základe obmedzení, ktoré definuje použitý algoritmus RELP vo Festivale a difónová metóda syntézy ako taká, by bolo zlepšenie daných hlasov možné najmä nahradením algoritmu RELP iným algoritmom, ktorý by nevyužíval LPC koeficienty, ale pre výber požadovaných difón by vstupoval priamo do akustickej databázy. Ďalšie zlepšenie týchto hlasov by mohlo byť dosiahnuté natrénovaním modelu prozódie na základe využitia štatistických informácií vyplývajúcich z rečovej databázy, ktorá by obsahovala nahrávky prirodzenej reči.

POĎAKOVANIE

Táto práca bola vytvorená realizáciou projektu Centrum informačných a komunikačných technológií pre znalostné systémy (kód ITMS projektu: 26220120030) na základe podpory operačného programu Výskum a vývoj financovaného z Európskeho fondu regionálneho rozvoja.

LITERATÚRA

[1] TAYLOR, P.: Text-to-Speech Synthesis. University of Cambridge, Cambridge : Cambridge University Press, 2009. 597 s. ISBN 978-0-521-89927-7
[2] BLACK, A. - TAYLOR, P. - CALEY, R.: The Festival Speech Synthesis System: System documentation. 1999.
[3] JUHÁR, J.: Rečové technológie v telekomunikačných a informačných systémoch. Košice : Equilibria, 2011. 519 s. ISBN 978-80-89284-75-7
[4] CERŇAK, M.: Využitie objektívnych meraní kvality pri korpusovej syntéze reči. Dizertačná práca. Slovenská technická univerzita v Bratislave, 2004. 119 s.
[5] HUANG, X. - ACERO, A. - HON, H.: Spoken Language Processing: A Guide to Theory, Algorithm and System Development. Carnegie Mellon University Pittsburgh : Prentice Hall, 2001. 1008 s. ISBN 0-13-022616-5
[6] ITU–T P.862. Perceptual evaluation of speech quality (PESQ) : An objective method for end-to-end speech quality assessment of narrow-band telephone networks and spe-ech codecs. International telecommunication union, 2001. 21 s.
[7] ITU–T P.862.1. Mapping function for transforming P.862 raw result scores to MOS-LQO. International telecommunication union, 2003. 3 s.

Programový modul funkcie predpovede počasia pre interaktívny rečový komunikátor

Marcel STRUCKEL, Matúš PLEVA

Katedra elektroniky a multimediálnych telekomunikácií, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

marcel.struckel@student.tuke.sk, matus.pleva@tuke.sk

***Abstrakt* — Táto práca sa týka problematiky získavania dát z webových zdrojov, ktoré slúžia pre rečový komunikátor na vývoj a testovanie rečových aplikácií v komunikácii človek-stroj. Získavanie dát z voľne dostupných zdrojov pozostáva väčšinou zo stiahnutia príslušnej webstránky, spracovania a následného odstránenia nepotrebných častí zdrojového kódu. Na ďalšie spracovanie sa v prípade služby Počasie používajú iba potrebné údaje o aktuálnom počasí, resp. údaje o predpovedi počasia na nasledujúce dni. Algoritmus je kombináciou skriptovacích jazykov a jazyka C, pričom modul je vytvorený ako súčasť Galaxy architektúry. Modul pracuje s údajmi, ktoré ďalej spracuje do požadovaného tvaru a pošle v štandardnom formáte Galaxy rámca rečovému komunikátoru, ktorý požadovanú predpoveď počasia prevedie z písomnej, do hovorovej formy. Výhodou použitia Galaxy infraštruktúry je, že modul môže byť jednoducho nahradený za iný bez zásahu do zdrojových kódov.**

***Kľúčové Slová* — html parsing, rečový komunikátor, Galaxy architektúra**

I. Úvod

Motiváciou vývoja nového modulu „Predpoveď počasia“ je úloha záložného systému pre momentálne dostupné a testované riešenia IRKR komunikátora [1]. Často sa totiž stáva, že sa stanú niektoré služby nefunkčné, pretože zdroj dát, ktorý využívali je dočasne nedostupný, zmenil štruktúru webu, či zmenil užívateľovi prístup k dostupným dátam. Systém, ktorý momentálne je využívaný hlavne komunitou zrakovo postihnutých poskytuje dôležitú pomôcku pre hendikepované osoby, a preto je snaha udržať ho vo funkčnom stave, a prípadne vylepšovať a testovať nové funkcionality v komunikácii človek stroj. Pôvodne bol systém určený len na jednoduchý dialóg s hlasovými povelmi s obmedzenou gramatikou [2], v súčasnosti už umožňuje zadanie požiadavky celou vetou s využitím rozpoznávača spojitej reči a modulu porozumenia prirodzenej reči (NLU – Natural Language Understanding) [3].

Vyvinutý modul „Predpoveď počasia“ cez interaktívny rečový komunikátor sa skladá z viacerých menších častí, ktoré budú postupne podrobnejšie opísané:

1. Zber dát a vytvorenie databázy s údajmi o počasí na nasledujúce dni pre 83 slovenských miest a obcí,
2. Parsovanie uložených html stránok – odstránenie redundancie, zapísanie potrebných údajov do samostatných súborov, podľa dní (pondelok.txt, utorok.txt, atď.).
3. Samostatný, neustále bežiaci modul ktorý si vymieňa informácie s rečovým komunikátorom, posiela mu požadovanú predpoveď počasia, pre konkrétny deň a mesto, vo forme Galaxy rámcov.

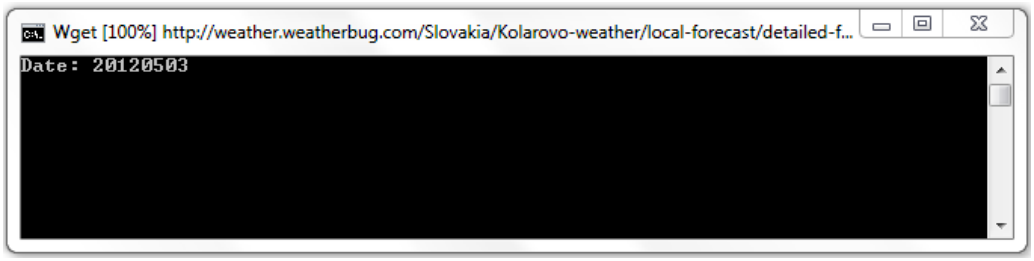
II. Zber dát potrebných pre postprocessing

Zdrojom údajov o počasí, s ktorými program pracuje, je zahraničná voľne dostupná webová stránka [4]. Práve táto webstránka bola zvolená z dôvodu, že sa na nej nachádzajú, a každodenne aktualizujú predpovede počasia na najbližších 7 dní v týždni, pre 83 Slovenských miest a obcí. Ďalším dôvodom výberu tejto stránky je ten, že nepodlieha častým zmenám grafického dizajnu, a celková štruktúra stránky sa nemení, čím sa zjednodušuje pravidelné prispôsobovanie zdrojového kódu parsera html stránok.

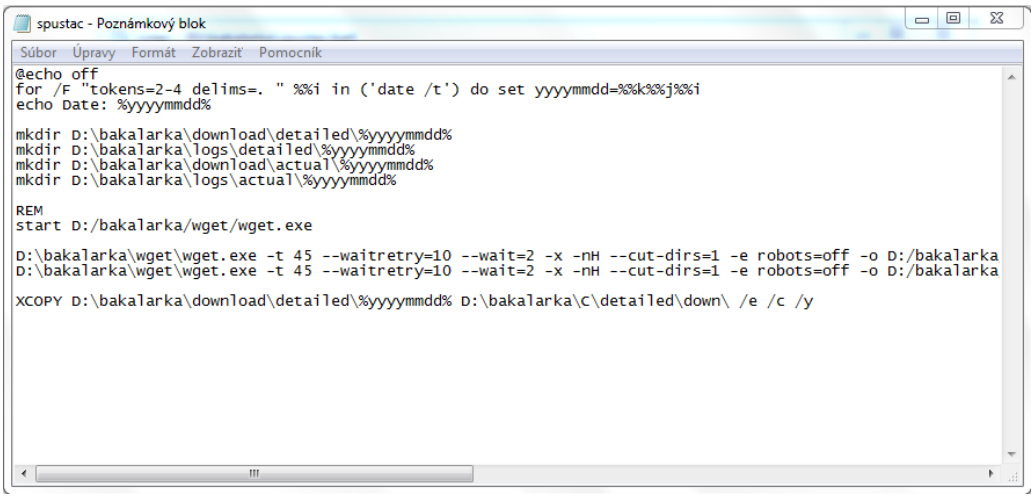
Na pravidelné sťahovanie a vytváranie databázy html stránok, s údajmi o počasí na nasledujúcich 7 dní, je použitý voľne dostupný program *Webget.exe* [5]. Úlohou tohto

programu je každodenné sťahovanie 83 html stránok s aktuálnymi predpoveďami počasia pre nasledujúce dni. Program *webget sa* spúšťa automaticky s využitím služieb operačného systému, každý deň o 02:00 hod. O spustenie, a správne fungovanie tohto programu sa stará spúšťací súbor *spustac.bat*, ktorý obsahuje konfiguráciu a inštrukcie pre program *webget*. V týchto inštrukciách sú zahrnuté webové stránky, ktorých obsah má byť stiahnutý, počet opakovaní sťahovania v prípade zlyhania prvého pokusu a cieľové adresy zložiek, do ktorých sa budú html stránky ukladať. Tieto zložky sú vytvárané každý deň, pričom ich názov zodpovedá aktuálnemu dátumu, pre jednoduchšiu orientáciu v databáze uložených html súborov. Ide teda o jednoduchú formu databázy, s využitým predefinovanej adresárovej štruktúry, ktorá slúži potom pre ďalšie časti modulu na získanie potrebnej informácie. Tento typ uloženia dát je veľmi efektívny, pretože nepotrebuje žiadnu ďalšiu réžiu a pre túto jednoduchú úlohu je prehľadný a rýchly.

Dodatkom k automatickému zberu dát je ďalší príkaz pre program webget, ktorý zaručí, že všetky aktuálne webstránky s údajmi o počasí sa uložia taktiež do samostatného súboru, z ktorého ich *parser* načíta pre svoje potreby. Tieto súbory sa každý deň prepisujú novými, čím vzniká vždy aktuálna databáza predpovedí počasia.



Obr. 1 Ukážka spusteného programu webget – vytvorenie zložky podľa aktuálneho dátumu a následné sťahovanie všetkých dostupných predpovedí počasia.



Obr. 2 súbor *spustac.bat* ktorý obsahuje inštrukcie na sťahovanie jednotlivých predpovedí pre program webget.

III. HTML Parser

Druhá časť celej funkcie predpovede počasia pre Slovenskú republiku, spočíva v parsovaní uložených html stránok, teda ide o "čistenie" daných html stránok od zbytočných, nepotrebných informácií.

O parsovanie webstránok a získavanie informácií o počasí na najbližších 7 dní, sa stará program vytvorený v jazyku C, *html_parser.exe.* Pre každý jeden html súbor (je ich 83), sa vytvorí textový súbor *parsed_názov mesta.txt*.

Program načítava znaky zo zdrojového súboru, teda uloženej webstránky, kontroluje každý jeden znak samostatne, a následne do vytvoreného textového súboru zapíše iba tie znaky, ktoré vyhodnotí ako potrebné. To sú všetky znaky, okrem html tagov ako napríklad <html>, </html>, <img src=…> atď., ďalej medzier, ukončení riadkov (predíde sa tým zbytočnému vytváraniu medzier a nových riadkov v textovom dokumente), adries a odkazov na iné webové stránky, skriptov a iných nevyužiteľných slov alebo znakov.

Na odstránenie klasických html tagov je použité rozlišovanie znaku, teda po načítaní jednotlivých znakov sa vždy urobí kontrola, či ide o znak <. Ak áno, znamená to že pointer vstupuje do html tagu, ktorého obsah nepotrebujeme prepisovať do textového súboru, a tak program načítava ďalšie znaky, až pokým nenačíta znak >. V tomto prípade je už pointer mimo html tagu, a ďalšie načítané znaky budú zaradené medzi potrebné, a prepísané do textového súboru pre dané mesto alebo obec.

IV. Parser Predpovedí

Ďalšou časťou celej funkcie predpovede počasia je opäť podprogram vytvorený v jazyku C, ktorého úlohou je získať, a zapísať do novovytvorených súborov, potrebné informácie z textového dokumentu ktorý vznikol prebehnutím predošlého podprogramu.

Tento parsovací program už pracuje na mierne vyššej úrovni, a to tým, že rozlišuje predpovede na všetkých sedem dní v týždni, pre každé mesto samostatne. V praxi to znamená, že z jedného textového súboru s informáciami, dokáže vytvoriť sedem nových textových súborov: *monday.txt, tuesday.txt, wednesday.txt, thursday.txt, friday.txt, saturday.txt* a *sunday.txt.* Po prebehnutí programu, každý z týchto súborov bude obsahovať predpoveď počasia pre daný deň, ako dennú, tak aj nočnú.

Každá z predpovedí obsahuje číselný údaj o maximálnej dennej teplote, rýchlosti vetra, vlhkosti, pocitovej a najnižšej nočnej teplote. V každej predpovedi je taktiež slovne vyjadrený stav počasia: *slnečno, polojasno, dážď, prehánky, sneženie, atď.*

A. Preklad získaných údajov do slovenčiny

Keďže dáta sú čerpané zo zahraničnej webstránky, všetky slovné údaje o stave počasia pre najbližších 7 dní sú v anglickom jazyku. Nakoľko funkcia predpovede počasia pre Slovenskú Republiku pracuje so slovenským jazykom, je potrebné preloženie potrebného textu z anglického, do slovenského jazyka. O preklad sa stará čiastočne funkcia slovníka.

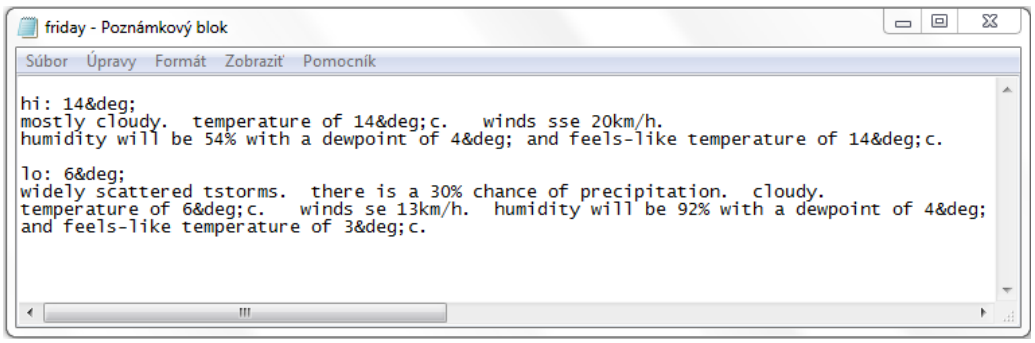
V textovom dokumente *preklad*.txt sú vypísané všetky anglické výrazy, ktoré sa môžu v predpovedi počasia vyskytnúť. Od každého takéhoto výrazu je tabelátorom (/t) oddelený výraz v slovenskom jazyku. Takéto rozloženie umožňuje prekladaču jednoducho porovnávať reťazce v predpovedi a v slovníku, a následne po nájdení zhody, je vybraný a použitý slovenský výraz ktorý je oddelený tabelátorom, resp. znakom */t*. Ide teda o jednoduchý tabuľkový prekladač.

Zvyšok prekladu je už pevne stanovený v poslednej časti programu, konkrétne vo výpise. Keďže štruktúra viet je v každej predpovedi rovnaká, výpis predpovede v slovenskom jazyku bude mať konkrétne stanovený tvar, a to:

[slovenský výraz zo slovníka], najvyššia denná teplota dosiahne %s stupňov Celzia, vietor s rýchlosťou %s kilometrov za hodinu, vlhkosť vzduchu dosiahne %s percent, pocitová teplota %s stupňov. V noci bude [slovenský výraz zo slovníka], najnižšia nočná teplota dosiahne %s stupňov Celzia, vietor s rýchlosťou %s kilometrov za hodinu, vlhkosť vzduchu dosiahne %s percent a pocitová teplota bude %s stupňov Celzia.

B. Výpis údajov

Samotný výpis je však doménou poslednej časti programu. Po prebehnutí tohto parsera, bude predpoveď rozdelená do samostatných súborov podľa príslušného dňa, čo umožňuje ďalšiemu programu jednoduchú orientáciu v priečinkoch a súboroch.



Obr. 3 Výsledok parsovania, ukážka predpovede na piatok, pre mesto Banská Bystrica.

V. VÝBER PREDPOVEDE

Predposlednou časťou modulu Predpovede počasia, je program *final.exe* vytvorený v jazyku C. Tento program priamo komunikuje s modulom, ktorý je súčasťou Galaxy architektúry [6] a komunikuje s IRKR rečovým komunikátorom, a teda dostáva požiadavky vo forme názvu mesta, pre ktoré je žiadaná predpoveď počasia, a následne prichádza požiadavka na konkrétny deň. Ako už bolo uvedené vyššie, dostupných miest a obcí je 83, a dni pre ktoré dokáže program určiť správnu predpoveď, môžu byť: pondelok, utorok, streda, štvrtok, piatok, sobota, nedeľa, dnes, zajtra, pozajtra, popozajtra, actual.

A. Aktuálny stav počasia

Pre aktuálny stav počasia, je vytvorený samostatný program *actual.exe*, ktorý funguje podobným spôsobom ako predchádzajúce programy na určenie predpovede počasia na nasledujúce dni. Keď od volajúceho príde cez interaktívny rečový komunikátor požiadavka na aktuálny stav počasia, spustí sa program *actual.exe*, ktorý najprv spôsobí spustenie programu *webget.exe*, cez ktorý sa stiahne a uloží aktuálny stav počasia z webstránky *weatherbug.com [4]*. Následne sa parsovaním vyberú iba číselné údaje, ku ktorým sa vo výpise priradí stav počasia v textovej forme v slovenskom jazyku. Konečný výstup aktuálneho počasia má teda tvar:
Aktuálna teplota: %s stupňov Celzia, Vietor: %s km/h, najvyššia očakávaná teplota: %s stupňov Celzia, najnižšia očakávaná teplota: %s stupňov Celzia.

B. Konverzia relatívneho dátumu na deň v týždni

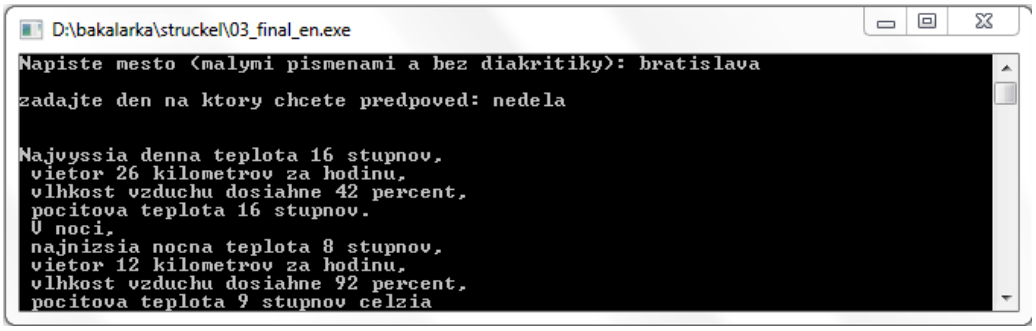
Aby program vedel slovnej požiadavke dnes, zajtra, pozajtra a popozajtra priradiť konkrétnu predpoveď, bolo potrebné implementovanie funkcie zisťovania času a dátumu, do programu *final.exe*. Táto funkcia pracuje s časovými knižnicami *time.h*. Každému dňu priradí číslo 0 – 6 (ned – 0, pon – 1, uto – 2, str – 3, štv – 4, pia – 5, sob – 6). Výstupom tejto časovej funkcie je teda číslo. Program *final.exe* podľa tohto čísla zistí aký je práve deň, a následne premenným dnes, zajtra, pozajtra a popozajtra priradí príslušný súbor s predpoveďou počasia. Teda ak je dnes napríklad *streda* (3), premenná *dnes* bude obsahovať cestu k súboru *wednesday.txt*, premenná *zajtra*, cestu k súboru *thursday.txt*, premenná *pozajtra*, cestu k súboru *friday.txt* a premenná *popozajtra* bude obsahovať cestu k súboru *saturday.txt*. Takto dokáže program priradiť cesty ku jednotlivým predpovediam pre každé jedno mesto a obec.

C. Výber údajov

Celý proces výberu a výpisu správnej predpovede teda prebieha nasledovne.

Ako prvá príde programu požiadavka na dané mesto alebo obec. Program overuje požiadavku so svojou databázou, a keď nájde zhodu, vnorí sa do zložky, ktorá nesie názov podľa požadovaného mesta alebo obce. V tejto zložke sa nachádza 7 textových súborov, s predpoveďami počasia pre najbližších sedem dní. Z toho vyplýva, že program potrebuje získať údaj o tom, ktorý z týchto siedmych súborov sa má otvoriť a načítať.

Ako už bolo vyššie popísané, program si vyberie z daného súboru iba prvý výraz, ktorý pomocou slovníka preloží do slovenčiny, a ďalej si načíta všetky číselné hodnoty do premenných. Následne vypíše požadovanú predpoveď v slovenskom jazyku, ktorú načíta do premennej ako reťazec znakov, a odošle rečovému komunikátoru, ktorý dokáže takýto text nahlas prečítať [1].

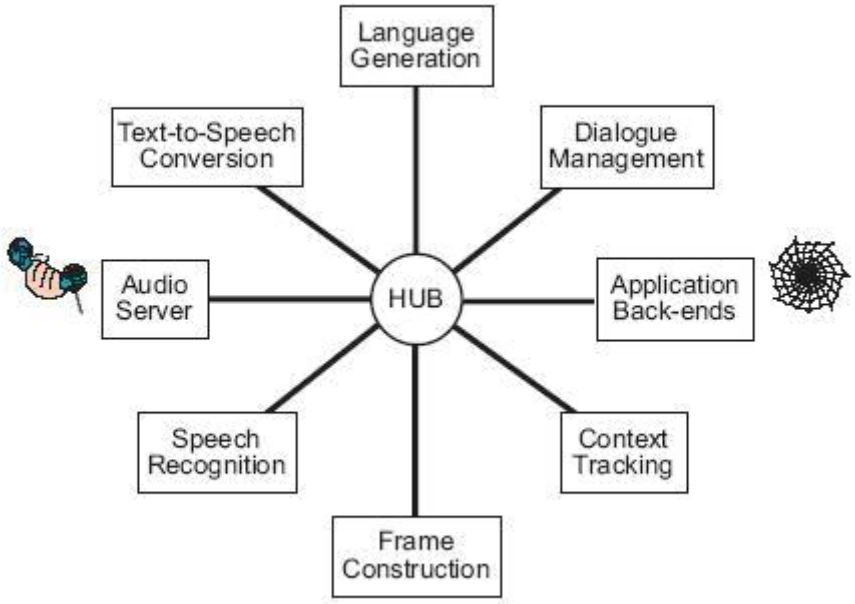


Obr. 4 Textový výpis predpovede počasia pre mesto Bratislava a deň nedeľa.

VI. GALAXY COMMUNICATOR A IRKR BACKEND MODUL

Inteligentné rečové komunikačné rozhranie (IRKR) je systém, ktorý umožňuje prístup k rôznym funkciám a informáciám (predpoveď počasia, cestovné poriadky) prostredníctvom telefónu, s možnosťou pripojenia do telekomunikačnej siete iba s využitím hlasového módu. Požadované informácie sa získavajú z lokálnych alebo externých dátových zdrojov. Princíp komunikácie je založený na technológii rozpoznávania a syntézy reči v slovenskom jazyku [7].

Základom IRKR komunikátora je Galaxy infraštruktúra, na ktorej sa nachádzajú servery komunikátora. Galaxy Communicator je systém založený na posielaní správ, optimalizovaný na vytváranie systémov dialógu. Aby bol modul „Počasie“ zabudovateľný do Galaxy komunikátora, musel byť vytvorený Galaxy modul nakonfigurovaný na funkciu „Backend“ a byť dodržaný formát správ, ktorý IRKR komunikátor používa pri požiadavkách na predpoveď počasia na modul získavania údajov (backend).



Obr. 5 Galaxy architektúra

Galaxy komunikátor je založený na HUB architektúre [8]. Využíva centrálny HUB proces, ktorý komunikuje na báze TCP/IP spojenia so všetkými modulmi s funkciami a podľa programu, ktorý má vo forme hub skriptu vopred zadefinovaný. Výhodou takejto komunikácie je, že každý komponent, či už HUB alebo server, môže byť spustený na rôznych PC pripojených do siete Internet a takisto môže fungovať na rôznych platformách (Windows, Linux, Solaris, ...), čo sa aj reálne využíva, pretože Audio server pre komunikáciu cez VoIP je postavený na Linuxe a beží spolu s Asterisk serverom cez ktorý vytvára hlasové spojenia. Všetky ostatné moduly sú spustené na Windows OS ale umožňujú využitie viacerých PC na distribúciu záťaže – hlavne rozpoznávania reči.

Backend modul komunikuje podľa nasledujúceho algoritmu: keď server prijme rámec s požadovanými informáciami (InFrame) od HUB-u, program *final.exe* rozlíši o akú obec a o ktorý deň sa jedná, a vytvorí rámec do ktorého sa načíta adekvátna odpoveď vo forme slovnej predpovede počasia. Takýto rámec, nazývaný *OutFrame,* sa odošle do systému IRKR (do HUB-u), ktorý ho následne podľa vlastných algoritmov prevedie z textovej do hovorovej formy, a teda volajúcemu sa dostane hovorená odpoveď na jeho požiadavku. Backend modul je neustále spustený, čo umožňuje IRKR odbavovať viacerých volajúcich aj v krátkom čase za sebou.

VII. ZÁVER

Inteligentné rečové komunikačné rozhranie je systém, ktorý poskytuje širokú škálu využitia. Vďaka systému IRKR, je možno aplikovať a zavádzať stále viac rôznych telefónnych služieb. V tomto prípade ide o službu predpovede počasia pre slovenské mestá a obce, ktorú možno využívať kdekoľvek na Slovensku. Na zistenie predpovede počasia kedykoľvek, 24 hodín denne, úplne postačuje telefónny aparát, a vedomosť o názve obce pre ktorú si účastník žiada informácie o počasí. Tieto informácie sa každý deň obnovujú, a teda sú vždy „čerstvé“. Pomocou funkcie pre predpoveď počasia, a systému IRKR, si môžu ľudia naplánovať výlety alebo pracovné stretnutia na nasledujúcich sedem dní. Okrem predpovede počasia umožňuje daný program v spolupráci s rečovým komunikátorom získať taktiež informáciu o vlakovom a autobusovom spojení, či informáciu o MHD v Košiciach.

Ako už bolo spomenuté vyššie, systém IRKR vytvára priestor pre tvorbu nových, prepracovaných služieb, ktoré dokážu človeku zjednodušiť vyhľadávanie informácii na internete. Takýmito službami môžu byť napríklad informácie o letových poriadkoch, o plánovaných koncertoch a festivaloch ako aj informácie o rôznych spoločnostiach či službách (pizzerie, kaviarne, reštaurácie, atď.). Je samozrejmosťou, že tento nezávislý systém má vysoký potenciál, a v budúcnosti bude ešte viac využívaný.

POĎAKOVANIE

Táto práca bola vytvorená realizáciou projektu Rozvoj Centra informačných a komunikačných technológií pre znalostné systémy (kód ITMS projektu: 26220120030) na základe podpory operačného programu Výskum a vývoj financovaného z Európskeho fondu regionálneho rozvoja.

LITERATÚRA

[1] http://irkr.tuke.sk/ - webová stránka projektu IRKR aj so službou monitorovania priebehu dialógu
[2] Juhár, J., Čižmár, A., Rusko, M., Trnka, M., Rozinaj, G., Jarina, R. “Voice operated information system in Slovak” (2007) Computing and Informatics, 26 (6), pp. 577-603.
[3] Pleva, M., Ondas, S., Juhar, J., Cizmar, A., Papaj, J., Dobos, L. “Speech and mobile technologies for cognitive communication and information systems” (2011) 2nd International Conference on Cognitive Infocommunications, CogInfoCom 2011, art. no. 5999458.
[4] http://weather.weatherbug.com/Slovakia-weather.html - použitý zdroj informácií o počasí
[5] http://www.gnu.org/software/wget/ - informácie o produkte wget (webget) a link na stiahnutie aj pre WinOS
[6] Juhar, J., Ondas, S., Cizmar, A., Rusko, M., Rozinaj, G., Jarina, R. “Development of Slovak GALAXY/VoiceXML based spoken language dialogue system to retrieve information from the internet” (2006) INTERSPEECH 2006 and 9th International Conference on Spoken Language Processing, INTERSPEECH 2006 - ICSLP, pp. 485-488.
[7] http://www.ktl.elf.stuba.sk/projects/speech/index.php?page=IRKR/IRKR
[8] http://communicator.sourceforge.net/sites/MITRE/distributions/GalaxyCommunicator/docs/manual/index.html

Systému na zber a analýzu analógových fyziologických signálov snímaných z mozgu a miechy pokusných zvierat

[1]Martin SZARKA, [1]Ján ŠALIGA, [2]Ján GÁLIK

[1]Katedra elektroniky a multimediálnych telekomunikácií, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika
[2]Úsav neurobiológie SAV, Košice, Slovenská republika

[1]szarka.martin@gmail.com, [1]jan.saliga@tuke.sk, [2]galik@saske.sk

***Abstrakt* — Tento článok prezentuje multikanálový systém na digitalizáciu elektrofyziologických signálov. Systém bol vytvorený na báze softvérového vybavenia v programovacom prostredí LabVIEW a multifunkčnej meracej karty. Systém umožňuje priame snímanie elektrofyziologických signálov z pokusného zvieraťa v reálnom čase so synchronizáciou na stimulovanie zvieraťa. Digitalizované dáta sú priebežne zobrazované ako aj ukladané do databázy v riadiacom počítači. Systém umožňuje aj následné vyberanie a analýzu signálov z tejto databázy.**

***Kľúčové slová* — meranie signálov, elektrofyziologické dáta, LabVIEW, evokované potenciály, systémy zberu dát**

I. Úvod

Jednou z možností, ako odhaľovať poruchy nervovej sústavy, je využitie výpočtovej techniky. V súčasnosti existuje veľká škála sofistikovaných komerčných meracích nástrojov, pomocou ktorých je možný záznam a analýza niektorých biologických signálov.

Cieľom tejto práce bolo vytvoriť systém, ktorý by pre svoju funkčnosť využíval čo najmenej ekonomicky náročný produkt a ktorý by slúžil Neurobiologickému ústavu pre špecifické merania. Pre danú úlohu sa javilo ako najviac vhodné použiť programovacie prostredie LabVIEW a jednoduché multifunkčné meracie karty od firmy National Instruments, keďže pomocou týchto nástrojov je možné jednoducho vytvoriť systém zberu dát, ktorý by spĺňal všetky požadované kritéria. Ďalšou výhodou použitia LabVIEW je jednoduchosť modifikácie programu, prípadne pridávanie ďalšej a rozširovanie stávajúcej funkcionality.

II. Fyziologické signály

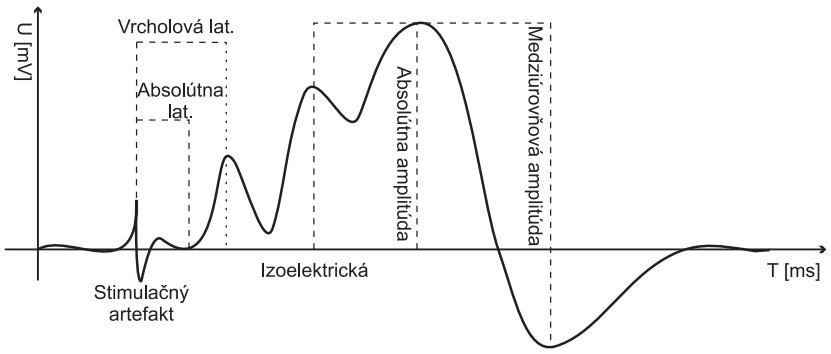
A. Evokované potenciály (EPs)

Evokované potenciály (EPs), taktiež nazývané vyvolané potenciály, sú odpoveďou nasledujúcou po predchádzajúcom externom stimule. Tieto potenciály sú využívané na vyšetrovanie funkcií a integrity nervových štruktúr. Takouto stimuláciou môžu byť elektrické impulzy, svetlo, grafické obrazce, zvuky, obrázky, slová a podobne. Odpoveď môže byť monitorovaná na ktoromkoľvek mieste pozdĺž dráhy. Môže napríklad ísť o odpoveď motorických vlákien v dolnej časti miechy na stimuláciu hornej časti miechy alebo môže byť monitorovaná odpoveď svalu na stimuláciu motorickej kôry mozgu [1].

V porovnaní s inými technikami je registrácia a následná interpretácia vyvolaných potenciálov jednoduchšia, pretože ich vyvolanie je priamo závislé od externe vyvolaného stimulu. Vyvolané potenciály sa vďaka jednoduchosti a vysokej informačnej hodnote stali dôležitou elektrofyziologickou metódou snímania fyziologických signálov [1].

B. Povaha evokovaných potenciálov (EPs)

EPs predstavujú zmenu elektrického potenciálu v snímanej oblasti. Takto nasnímaný potenciál sa vyznačuje niekoľkými elektrofyziologickými parametrami **(Obr. 1)**.



Obr. 1 Parametre EPs vlny [1].

Polarita je základným parametrom EP, pričom môže ísť o pozitívne alebo negatívne vlny EPs [3]. V prípade monopolárneho snímania polarita vlny korešponduje so skutočným napätím a to bez ohľadu na to, či ide o pozitívnu, alebo negatívnu hodnotu, ktorá je detegovaná aktívnymi elektródami [3]. Pri použití bipolárneho snímania pomocou diferenčného zosilňovača je polarita relatívna a závisí na tom, ktorá elektróda je v danom momente pozitívna, resp. negatívna vzhľadom na druhú [3].
Amplitúda predstavuje výšku alebo veľkosť EP vo voltoch. Meranie takejto amplitúdy je značne problematické, pretože je často nemožné znázorniť pokojovú líniu snímaného signálu. V prípade, že nie je možné zmerať skutočnú amplitúdu, je jednou možnosťou zmeranie amplitúd medzi vrcholmi susedných vĺn (medzivrcholová amplitúda). Ak sa tieto vlny čiastočne prekrývajú alebo je signál tvorený niekoľkými po sebe idúcimi vlnami, nastáva problém so zadefinovaním správnej skutočnej amplitúdy danej vlny. [2]
Latencia resp. skutočná vlnová latencia predstavuje čas, ktorý uplynie od momentu simulácie po zaznamenanie prvej vlny EPs. V niektorých prípadoch, podobne ako pri určovaní amplitúdy, je veľmi obtiažne určiť začiatok danej vlny, teda nájsť pokojovú líniu [2]. Preto sa v praxi určuje vrcholová latencia (peak latency), ktorá predstavuje časový interval od vysielania stimulu až po vrchol prvej EPs vlny [2].

C. Eliminácia rušivých činiteľov počas merania

Najväčším problémom snímania nervových vzruchov je nízky odstup signálu od šumu. Dôvodom je to, že amplitúda snímaných signálov sa často pohybuje rádovo od μV po mV. Amplitúda šumu má podobné hodnoty. Šum do merania môžu vnášať aj elektrické zariadenia, ktoré sú v blízkosti snímania.

Azda najefektívnejší spôsob eliminovania vonkajších vplyvov na snímanie je použitie Faradayovej klietky počas merania. Pomocou nej je možné vylúčiť vplyv okolitého elektromagnetického smogu na snímaný signál. Ďalšou možnosťou je použitie batérii pre napájanie prístrojov, aby sa zamedzilo rušeniu pochádzajúcemu z rozvodnej elektrickej siete. Šum sa dá redukovať aj použitím bipolárnej metódy snímania, ak to situácia umožňuje. Filtráciu šumu je vhodné použiť, ak sa frekvenčný rozsah podstatne líši od frekvenčného rozsahu šumu. Pre dosiahnutie optimálnych výsledkov je nutné poznať charakter snímaného signálu. Navrhovaný program bude šum eliminovať spriemerňovaním viacerých meraní čím sa zvýši odstup signálu od šumu, keďže šum je nekorelovaný.

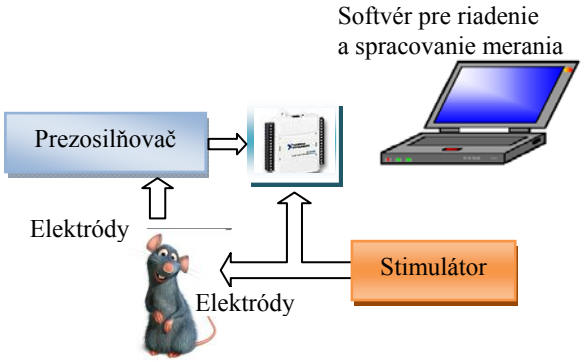
III. NÁVRH SYSTÉMU NA SNÍMANIE FYZIOLOGICKÝCH SIGNÁLOV

Pracovisko pre meranie EP je zobrazené na obr. 2 a 3. Programové vybavenie navrhnutého systému bolo rozdelené do dvoch hlavných častí a to časť, v ktorej sa signál synchronizovane so stimulmi meral, spriemerňoval a filtroval a následne ukladal do súboru v databáze. Druhá, samostatná časť umožňuje načítavanie merania zo súboru s následnou vizualizáciou a aplikovaním číslicového spracovania signálu pre potlačenie šumu a nežiadúcich zložiek v meranom signáli.

A. Snímacia časť

Program realizuje snímanie tak, že na digitálnom výstupe meracej karty generuje impulz, ktorý slúži ako pre spúšťanie stimulátora . Tento impulz je privádzaný samostatným kanálom spätne na vstup meracej karty pre potreby synchronizácie meraní v čase. Takáto synchronizácia je potrebná pretože generovanie spúšťacích impulzov sú ovplyvňované časovaním operačného systému na ktorom sa program spúšťa. Užívateľ má v programe možnosť nastavovať parametre synchronizácie (rozhodovacia úroveň napätia, pretrigering), parametre merania (vzorkovacia frekvencia, dĺžka snímania, frekvencia opakovania meraní, počet meraní pre výpočet

spriemernenia). Za synchronizačnú udalosť sa považuje prekročenie rozhodovacej úrovne v kanále, ktorý sníma generovaný stimulačný impulz. Program umožňuje automatické ukladanie merania vo formáte textového dokumentu (*.txt), binárneho súboru (*.dat), alebo binárneho formátu TDMS (*.tdms), ktorý je možné následne jednoducho načítať v programe MS Excel po doinštalovaní podpory pre TDMS súbory.



Obr. 2. Zostava experimentálneho pracoviska.



Obr. 3. Pohľad na reálne experimentálne pracovisko.

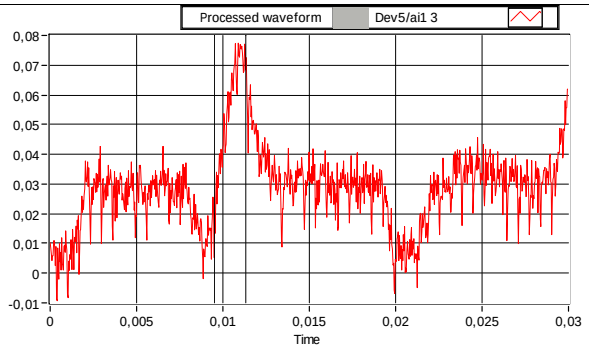
B. Načítavanie dát

Časť programu pre off-line spracovanie a vyhodnotenie uložených dát umožňuje načítavanie uložených dát, pričom súbor môže obsahovať dáta koncentrované z viacerých kanálov. Program preto umožňuje výber kanálov ktoré sa majú zobraziť. Je možné zvoliť režim, ktorý neprepisuje už zobrazené dáta. Tento režim umožňuje porovnať signály z viacerých súborov. Podmienkou takéhoto zobrazenia je to, že všetky porovnávané signály boli snímané rovnakou vzorkovacou frekvenciou. Nesplnenie tejto podmienky vedie k chybovému hláseniu a upozorneniu na nemožnosť vykonať požadovanú operáciu pre užívateľa. Zobrazenie grafu sa v takomto prípade nezmení a je tak možné pridať iný súbor, ktorý má rovnakú vzorkovaciu frekvenciu ako už zobrazené dáta. Ak sa pridávané dáta líšia v nastavení pretrigeringu, tak je vykonaná korekcia zobrazenia tak, aby boli zobrazené všetky dáta. Použije sa vždy väčšia z hodnôt pretrigeringu. Korekcia sa vykonáva vložením nulového poľa. Podobná korekcia sa vykonáva, ak je jeden zo zobrazovaných signálov kratší ako druhy. V takomto prípade sa na koniec kratšieho poľa (reprezentujúceho signál) pridajú nuly.

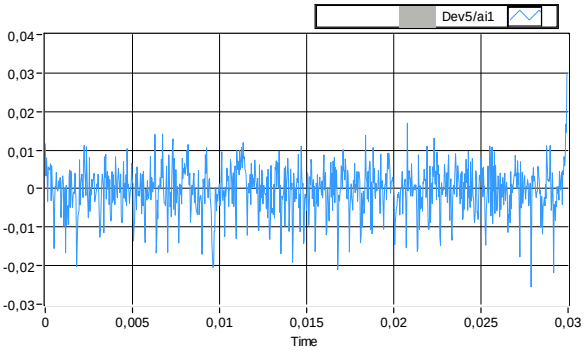
Táto základná verzia vyhodnocovacieho programu bola doplnená a rozšírená o vybrané pokročilé metódy číslicového spracovania signálu, ktorá umožňuje použiť rôzne nástroje a postupy pre potlačenie šumu a rušenia v meraných signáloch (nelineárna a lineárna filtrácia) ako aj detekciu signálu skrytého v šume s rušení pomocou Short time FFT.

Ako príklad aplikácie a účinnosti týchto metód sú na obrázkoch 4. a 5. zobrazené priebehy pred filtráciou a detekciou s dobrým a zlým pomerom signál šum. Na obrázku 6. Je zobrazený pôvodný zašumený signál spolu s výsledkom po filtrácii nelineárnym mediánovým filtrom a následne Besselovým IIR filtrom.

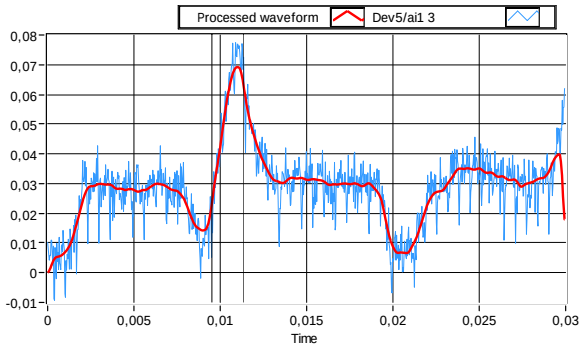
Obrázok 7 ukazuje výsledok ST FFT aplikovanej na priebeh z obrázku 5, ktorý naznačuje prítomnosť hľadaného signálu v čase okolo 10ms. Aplikovaním .



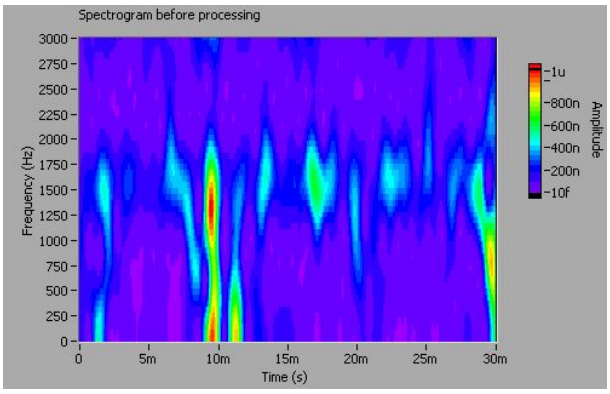
Obr. 4. Meraný signál s dobrým pomerom signál šum



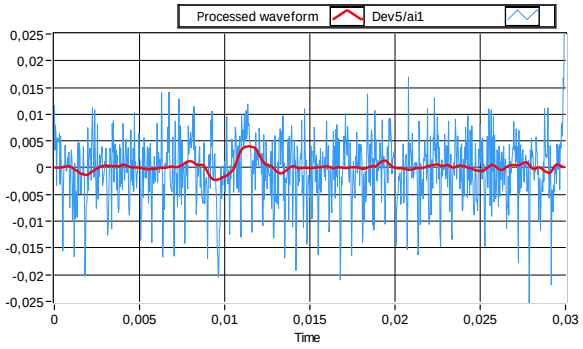
Obr. 5. Meraný signál s „stratený“ v šume



Obr. 6. Pôvodný signál z obr. 4 a výsledok mediánovej nelineárnej filtrácie a následnou aplikáciou IIR filtra



Obr. 7. Detekcia prítomnosti signálu prekrytého šumom pomocou ST FFT



Obr. 8. Pôvodný signál z obr. 5, detekovaný na obr. 7 a výsledok mediánovej nelineárnej filtrácie a následnou aplikáciou IIR filtra

IV. ZÁVER

Počas vývoja systému boli postupne dopĺňané ďalšie funkcie predovšetkým v oblasti pokrokových metód spracovania meraných signálov a v tomto procese sa pokračuje aj naďalej. Predpokladá sa implementovania ďalších pokročilých metód spracovania signálu ako je wavelet filtrácia s cieľom ešte lepšie potlačiť nežiaduce zložky v signáli a rekonštruovať pôvodný signál čo možno najpresnejšie.

V. POĎAKOVANIE

Táto práca bola vytvorená realizáciou projektu Rozvoj Centra informačných a komunikačných technológií pre znalostné systémy (kód ITMS projektu: 26220120030) na základe podpory operačného programu Výskum a vývoj financovaného z Európskeho fondu regionálneho rozvoja.

LITERATÚRA

[1] GÁLIK, Ján – CONWAY, CH.M. 1997. Evoked potentials: Principles and techniques. In Kopf Carrier. ISSN 1083-7396, 1997.
[2] CIGÁNEK, Leodegar: Evokované potenciály a ich využitie v klinickej praxi. Martin, Osveta, 1991.
[3] GÁLIK, Ján: Aplikácia niektorých elektrofyziologických vyšetrovacích metód pri ischemických a traumatických poruchách CNS. Kandidátska dizertačná práca, NBU SAV, Košice, 1995.

Comparison of maximum likelihood and least squares fitting for ADC testing in simulated environment

[1]Ján ŠALIGA, [1]Linus MICHAELI, [2]Ján BUŠA, [1]Jozef LIPTÁK, [1]Marek GODLA

[1]Department of Electronics and Multimedia Communications, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Slovak Republic
[2]Department of Mathematics and Theoretical Informatics, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Slovak Republic

[1]jan.saliga@tuke.sk, [1]linus.michaeli@tuke.sk, [2]jan.busa@tuke.sk,
[1]jozef.liptak@tuke.sk, [1]marek.godla@tuke.sk

***Abstract* — ADC test methods require the best possible reconstruction of ADC under test input signal from acquired, therefore erroneous ADC output data. The commonly used LS fit and the recently introduced maximum likelihood estimation are the competing methods. This paper presents an experimental comparative study of these estimation methods with the goal to check deeper the behavior of both methods and to determine their limits. A new algorithm for calculation of maximum likelihood fit is also examined. Authors will give some practical recommendation for choice of optimal method for various condition of ADC testing as a conclusion of the comparative study.**

***Keywords* — ADC test, maximum likelihood, least mean squares method, LS method, four-parameter fit, signal recovery, estimation of signal parameters**

I. INTRODUCTION

Standardized dynamic test methods for analog-to-digital converter (ADC) ([1] – [3]) are based on comparison of acquired ADC output codes with ADC input stimulus. The stimulus is not exactly known and therefore it must be reconstructed from the erroneous ADC output codes acquired during testing. Any inaccuracy in estimation of ADC stimulus parameters leads to inaccuracy in determination of ADC parameters measured by dynamic test.

The most common way how to recover input signal and estimate its parameters is least squares (LS) fitting. According to the theory the LS fitting gives the best estimation under the condition that the observation (quantization) noise is independent, white, and normally distributed with zero mean ([4], [5]). This is clearly not true for ADC testing and therefore the LS fit is usually worse than ML estimation would be – if one could obtain

n the latter using the properties of the quantization noise.

The more general way is fitting based on maximum likelihood (ML) method. This idea was introduced in [6] and [7] for sine wave and later it was generalized also for exponential stimulus ([8], [9]). The ML estimation is optimal is certain sense, but its main disadvantages are its computation complexity and convergence problems due to several local minima.

Although both fitting procedures are known, until now no deeper comparative study on limitation of methods has been performed. The main novelty of this paper is just this comparative research. Moreover, we also examined a new maximization method for ML fit. In contrary of the previous works ([6] – [9]) we performed calculations and signal processing in LabVIEW instead of Matlab.

II. FITTING

The general setup for dynamic ADC testing is shown in Fig. 1. To perform simulation experiments we developed software equivalent modules including non-ideal ADC model with optional test specific INL.

A. LS fit

LS fitting is well known and very often used procedure for recovering distorted and noisy data in testing and measurements. Estimation of parameters of ADC input signal is obtained from minimization of the cost function CF_{LS}:

$$\min\left(\mathrm{CF}_{\mathrm{LS}}(\mathbf{a})\right) = \min \sum_{n=0}^{N-1} \left(y(n) - \mathrm{f}(\mathbf{a}, n)\right)^2 \tag{1}$$

where f(**a**,*t*) is the time model of the testing signal, **a** is a vector of its unknown parameters, and $y(n)$ are digitized samples of testing signal applied to the input of ADC under test, taken at sampling instances nT_s (T_s is the sampling period). Solution of (1) leads to a system of equations which is often nonlinear, e.g., in the cases of four parameters fit of a sine wave or for dual slope exponential stimulus. Solution of such a nonlinear system requires application of an appropriate numerical method, and this increases the complexity of the LS method.

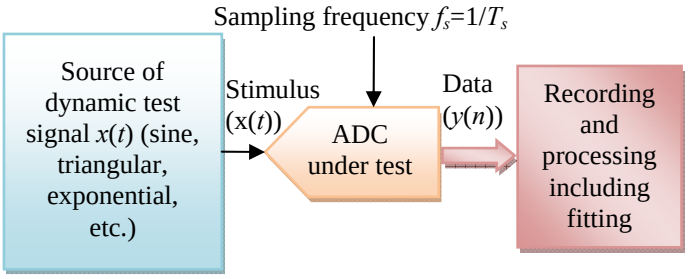


Fig. 1 General setup for dynamic ADC testing

B. Maximum likelihood fit

Maximum likelihood method comes from estimation theory and looks for the most probable ADC input signal. The ML fitting is based on maximization of a likelihood function $L(\mathbf{a})$:

$$\max_{\mathbf{a},\mathbf{q},\sigma}\left(L(\mathbf{a})\right) = \max_{\mathbf{a},\mathbf{q},\sigma} \prod_{n=0}^{N-1} P\left(y(n) = Y_{\mathbf{a},\mathbf{q},\sigma}(n)\right), \tag{2}$$

where **q** is the vector of ADC transition code levels, σ is the standard deviation of the Gaussian noise of the input electronics. $P(y(n) = Y(n))$ is the probability that the n-th ADC output sample $y(n)$ is equal to the expected output value $Y(n)$, which represents a possible output code. To simplify the maximization $L(\mathbf{a})$, minimization of the sum of negative logarithms of probabilities P(.) is preferred.

$$\arg\max_{\mathbf{a},\mathbf{q},\sigma}\left(L(\mathbf{a})\right) \approx \arg\min_{\mathbf{a},\mathbf{q},\sigma}\left(-\ln\left(L(\mathbf{a})\right)\right) = \arg\min_{\mathbf{a},\mathbf{q},\sigma}\left(-\sum_{n=0}^{N-1} \ln P\left(y(n) = Y_{\mathbf{a},\mathbf{q},\sigma}(n)\right)\right) \tag{3}$$

Searching the extreme is a complex task and it can be performed only by an appropriate numerical method.

C. Implementation of fitting and comparison

Experiments below were performed in software developed in LabVIEW. Properties, advantages and disadvantages of LabVIEW are generally well known. We choose LabVIEW not only because of our previous experience but also because we wanted to compare our results with previous experiments performed in Matlab ([5] – [8]).

LS fit for sine wave was implemented according to standards ([1] – [3]) in the form of 3 and 4 parameters fit. ML fit implementation was rather more complex because of the complexity of numerical minimization of negative logarithm of likelihood function. In contrast to minimization methods based on gradients used in [5] – [8], the LabVIEW built-in function

`GlobalOptimization.vi` ([9]) has been utilized. This function is based on the Differential Evolution method (DE), which is one of general methods used to solve the global optimization problem. DE methods approximate the global optimum by mutating and improving the candidate parameters from the initial ones. Fig. 2 illustrates how the algorithm works. GlobalOptimalization.vi uses (if available) multi-core processing, which speeds up the calculation of the best parameters.

Comparison of accuracy of signal recovering achieved by LS and LM methods utilizes two parameters:

1. Estimation error defined as RMS value of difference between recovered signal and original input signal.
2. Difference between SINADs calculated from LS or from ML fit and the exact value calculated from exactly know ADC stimulus.

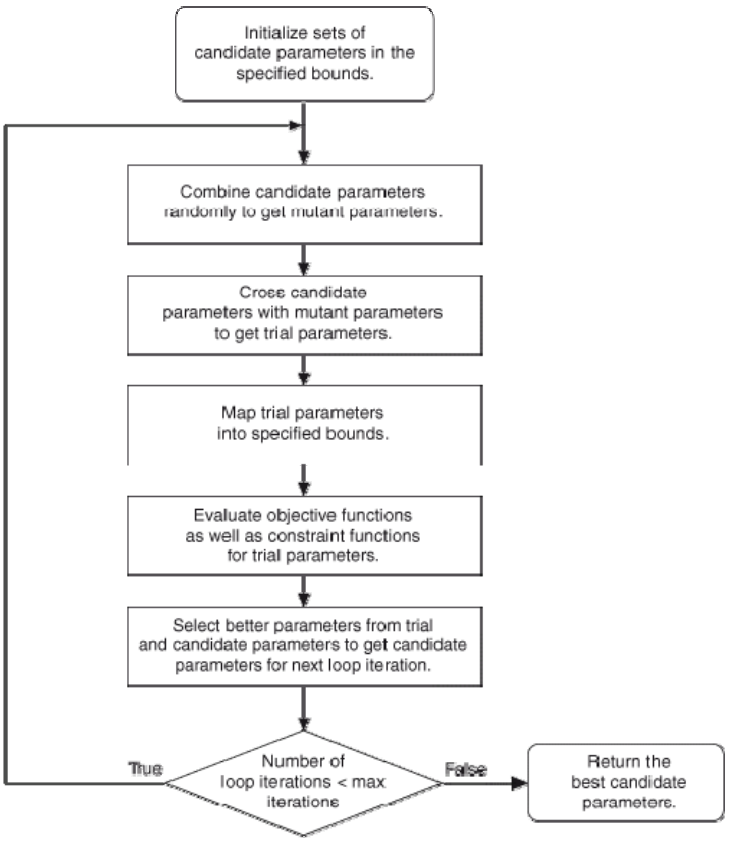


Fig. 2 Algorithm of DE method

RMS value (standard deviation) was chosen because it is the basic component, which has influence on nearly all final ADC dynamic parameters such as ENOB and SINAD. Moreover, a simple comparison of differences between, e.g., sine wave amplitudes is not well readable from the point view of its influence on the final error in calculation of ADC parameters. The estimation error - RMS value of the difference between the recovered signal by a given method D_{method} and the known exact one was calculated according to the formula:

$$D_{\text{method}} = \sqrt{\frac{1}{N}\sum_{n=1}^{N-1}\left(y_{\text{method}}(n) - y_{\text{exact}}(n)\right)^2}\,, \tag{4}$$

where $y_{\text{method}}(n)$ are the samples calculated from signal recovered by the giving method and $y_{\text{exact}}(n)$ are the equivalent samples of the exact ADC stimulus.

Comparison of SINAD calculated from LS and ML fits to SINAD calculated from the exact value of stimulus was utilized only in tests, where the stimulating sine wave did not overload ADC full scale as it is required by standards. SINAD was calculated according to the standard defined formula:

$$SINAD_{\text{method}} = 20\log\frac{A_{\text{method}}}{D_{\text{method}}}\,, \tag{5}$$

where A_{method} is RMS value of sine wave recovered by a given method or from exactly known ADC stimulus.

III. Experimental Results

The first experiments were performed on linear ideal ADC with user specifiable resolution (up to 16 bits), and standard deviation of noise (0.5LSB), which represented ADC code alternating behavior. Records with optional number of samples were utilized for fitting by a chosen method. Fig. 3 shows a zoomed segment of record of digitized stimulus from the ideal 8-bits ADC together with exact stimulus, and recovered signals. Stimulus did not overload ADC full scale. The differences between recovered signals and exact stimulus are hardly visible.

Fig. 4 shows the same signals on the condition that input signal overloads ADC full scale. In accordance with assumption ML fit give much better estimation of ADC stimulus than LS fit.

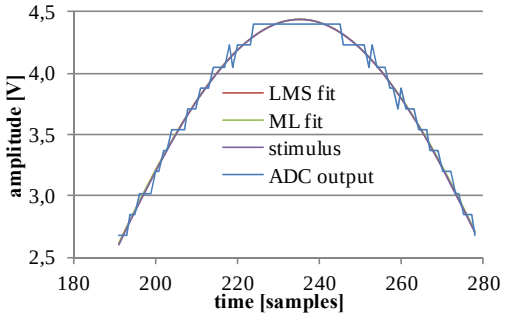


Fig. 3. Comparison of LS and ML fit, digitized and exact stimulus (zoomed time segment)

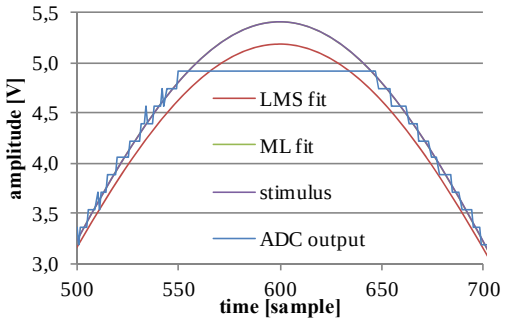


Fig. 4. Comparison of LS and ML fit, digitized and original stimulus (zoomed time segment) if stimulus overloads ADC full range. Both fits utilized all samples in record. LS

Next, the dependence of estimation error on signal covering of ADC input range was treated. Fig. 5 shows the dependence acquired for 8 bits ADC and 500 coherently taken samples with standard deviation of noise σ = 0.2LSB. If input signal covers ADC range up to 100 percent, both methods give nearly the same results. Very similar results were achieved for other ADC resolution, σ, and numbers of samples in record. The resulting practical recommendation is that the dominating factor for choice of fitting method is ADC overloading. ML fit is much better for any, even very small, ADC overloading.

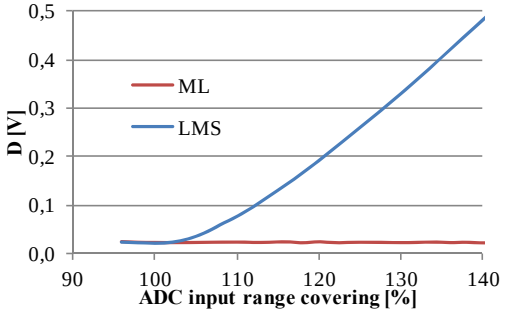


Fig. 5. LS vs. ML fit for overloaded ADC (100% is equal to ADC full scale). Both fits utilize all samples in record LS

Next study was focused on the accuracy of recovered stimulus for different resolutions of

ADC on the condition that ADC under test was not overloaded and σ = 1LSB and $N = 2^{b+1}$, where b is the ADC resolution in bits. The achieved results are shown in Figs. 6 and 7. The achieved results show that both methods give very similar and very good estimations. ML fit gives a little smaller estimation error (D) but the differences are in order of the variance of results in repeated experiments (see the variances in the figures).

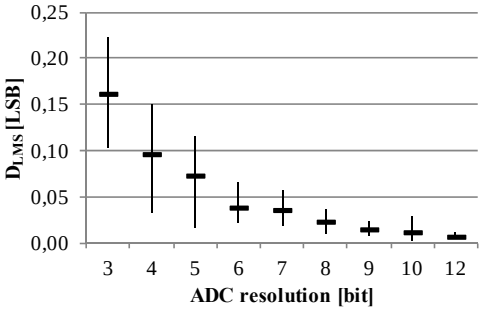


Fig. 6. Dependence of estimation error on number of ADC bits for LS fit

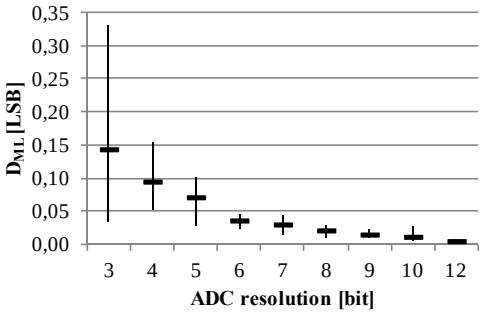


Fig. 7. Dependence of estimation error on number of ADC bits for ML fit

The next experiment was focused on dependence of estimation error D on number of samples in record at coherent sampling. Fig. 8 and Fig. 9 show this dependence for 8-bits ADC and σ = 0.5LSB. Both errors are very small and decrease with increasing number of samples. Again as in previous experiments, the variance of results is in the same order as mean value for repeated experiment. Fig. 10 presents the difference of errors' mean value between LS fit and ML fit.

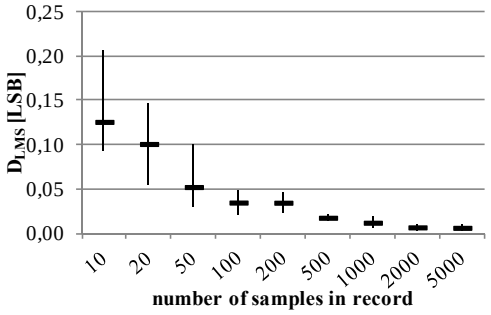


Fig. 8. Dependence of estimation error on number of samples for 8 bits ADC for LS fit

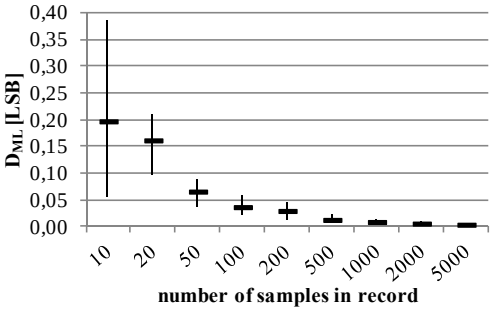


Fig. 9. Dependence of estimation error on number of samples for 8 bits ADC for ML fit

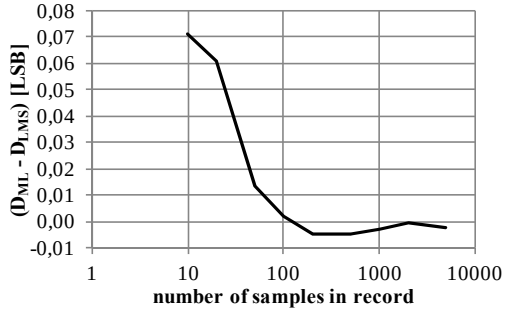


Fig. 10. Difference of error of ML and LS fits in dependence on number of samples in record

The next experiments were focused on influence of the ADC noise that causes alternating of ADC output codes. The results achieved for 8 bits ADC and 400 samples in record are shown in Fig. 11 and Fig. 12. ML fit is better for low noise and with increasing noise both fits give nearly the same estimation error. Variances of results for repeated experiments are also comparable with mean values of the error as it was in previous cases.

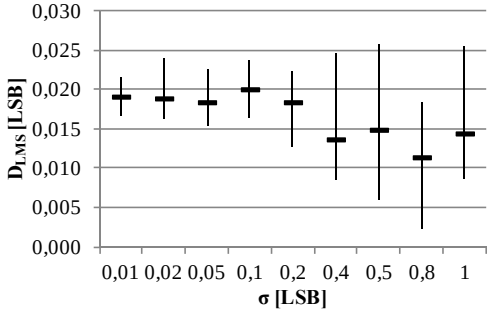


Fig. 11. Dependence of LS fit error on noise level

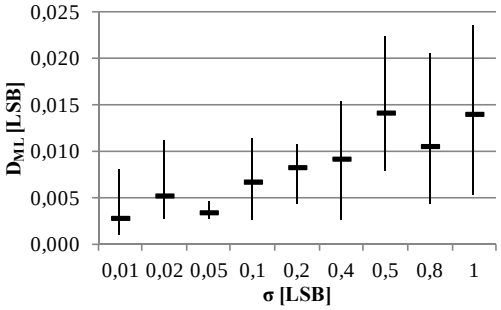


Fig. 12. Dependence of ML fit error on noise level

IV. Conclusions

This paper presented a comparative experimental study of LS and ML fit for ADC testing. The achieved results show that both methods give the very similar results for the lightly distorted signal and signal with low additional noise that is equivalent to oscillations of ADC at closing ADC input signal to a transition levels. Due to calculation complexity the LS may be preferred. If ADC is even slightly overloaded, the ML fit enables better testing signal especially for low resolution ADC (less than 10). Using the new numerical minimization methods the process of ML fit can be simplified if it is implemented in LabVIEW.

Our next work will bring more experimental results including SINAD test and more precise recommendations for application of LS and ML fit in ADC testing.

ACKNOWLEDGMENT

The work is a part of the project supported by the Science Grant Agency of the Slovak Republic (No. 1/0555/11).

This work was supported by the Agency of the Ministry of Education of the Slovak Republic for the Structural Funds of the EU under the project Development of Centre of Information and Communication Technologies for Knowledge Systems (project number: 26220120030).

REFERENCES

[1] IEEE Standard 1241-2000, "Standard for Terminology and Test Methods for Analog-to-Digital Converters", The Institute of Electrical and Electronics Engineers.

[2] IEEE Standard 1057-2007, "Standard for Digitizing Waveform Recorders", The Institute of Electrical and Electronics Engineers.

[3] D. Dallet and J. M. da Silva: "Dynamic Characterisation of Analogue-to-Digital converters." Springer 2005, ISBN-13 978-0-387-25902-4

[4] J. J. Blair: "Sine-fitting software for IEEE standards 1057 and 1241", in Proc. of the 16th IEEE Instrumentation and Measurement Technology Conference, IMTC/99, Venice, Italy, May 1999, vol. 3, pp. 1504-1506.

[5] I. Kollár and J. J. Blair: "Improved Determination of the Best Fitting Sine Wave in ADC Testing", IEEE Trans. on Instrumentation and Measurement, Vol. 54 No. 5, pp. 1978-1983, 2005.

[7] L. Balogh, B. Fodor, A. Sárhegyi, I. Kollár: "Maximum Likelihood Estimation of ADC Parameters from Sine Wave Test Data", In: Proc. of 15th IMEKO TC4 Symposium on Novelties in Electrical Measurements and Instrumentation: 12th Workshop on ADC Modelling and Testing. Iasi, Romania, 2007. (IMEKO) pp. 85-90.

[8] L Balogh, I Kollár, L Michaeli, J. Šaliga, J. Lipták: "Extracting Full Information from Measured ADC Data". In: 15th International Workshop on ADC Modelling and Testing. Kosice, Slovakia, 2010 (IMEKO) 6 p. Paper 121

[9] László Balogh, István Kollár, Linus Michaeli, Ján Šaliga, Jozef Lipták, Full Information from Measured ADC Test Data using Maximum Likelihood Estimation. MEASUREMENT 45:(2) pp. 164-169. (2012)

[10] LabVIEW 2011 built-in help/guide

Senzor hmly na LOS KEMT FEI TUKE

Matúš TATARKO, Ján TURÁN, Ľuboš OVSENÍK

Katedra elektroniky a multimediálnych telekomunikácií, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

matus.tatarko@tuke.sk, jan.turan@tuke.sk, lubos.ovsenik@tuke.sk

***Abstrakt* — V dnešnej modernej dobe sa zvyšuje požiadavka na nárast prenosových rýchlostí, ktorými sme schopní prenášať informácie koncovým užívateľom. Taktiež nadviazanie spojenia nebolo nikdy tak jednoduché ako dnes. Optické komunikačné systémy, ktorých prenosové prostredie je vzduch, sa stávajú čoraz populárnejšie na zabezpečenie takýchto požiadaviek. Použitím FSO (Free Space Optics) systémov, ako riešenie poslednej míle, je možné zabezpečiť spojenie za niekoľko hodín bez potreby drahej optickej kabeláže a licencovaných spektier. Medzi súčasné riešenia vďaka výkonným vysielacím laserom patrí aj umiestenie FSO systémov vo vnútri budov, pričom sa zvyšuje aj bezpečnosť týchto systémov, keďže sa zamedzí prístup k hlaviciam nepovolaným osobám. Najväčším „nepriateľom“ týchto systémov je nepriazeň počasia, ktorá vplýva na kvalitu prenosu a tým ovplyvňuje aj ich dostupnosť. Tento článok sa preto podrobne zaoberá popisom prístroja (Fog senzora) na meranie hustoty hmly, teploty a relatívnej vlhkosti prostredia. Dlhodobé merania týchto parametrov nám umožňujú vytvárať modely pre prenos signálu pomocou optických FSO systémov.**

***Kľúčové slová* — FSO systém, senzor hmly, hustota hmly, relatívna vlhkosť, teplota**

I. Úvod

FSO zariadenia používajú na komunikáciu voľným prostredím modulovaný viditeľný alebo neviditeľný lúč svetla [1-3]. FSO systém pozostáva z vysielača a prijímača, ktoré medzi sebou posielajú správy pomocou optických lúčov z laserových diód. Prijímač optického lúča je vybavený vysoko citlivým fotodetektorom. Tieto systémy pracujú na vzdialenosť niekoľkých kilometrov, ale musí byť zabezpečená priama viditeľnosť medzi vysielačom a prijímačom. Kvôli tomu, že tieto systémy používajú ako prenosové médium vzduch, existuje niekoľko typov rušiacich efektov v atmosfére, ktoré vplývajú na ich dostupnosť [1,3]. Hlavným rušiacim elementom optickej komunikácie voľným prostredím je hustá hmla. Hmla pozostáva z vodnej pary alebo vodných kvapôčiek, ktorých priemer je len niekoľko sto mikrometrov. Práve rozmer kvapôčiek vody má rušivý vplyv na optický signál s vlnovými dĺžkami, ktoré sa používajú v týchto systémoch (850 nm, 1550 nm). Dochádza k útlmom a rozptylom optického lúča, čím sa zníži prenášaný výkon optického lúča a následne nastane prerušenie komunikácie. Preto je potrebné urobiť predbežnú analýzu optického prenosu a na základe tejto analýzy stanoviť vhodné parametre prenosového systému. Jeden z prístrojov, ktorý dokáže merať kvalitu prenosového prostredia je náš senzor hmly (Fog senzor).

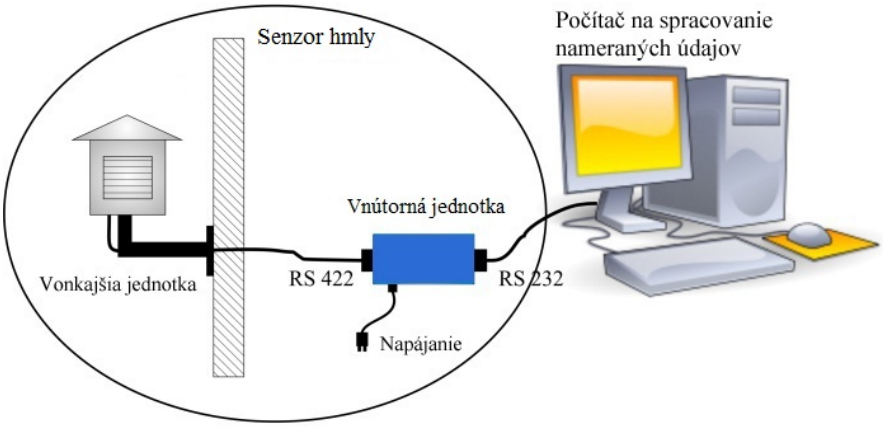
V prvej časti sú popísané všeobecné princípy toho zariadenia a jeho technické časti. V nasledujúcej kapitole popisujeme experimentálne výsledky a merania hustoty hmly.

II. Senzor hmly

Senzor hmly je zariadenie, ktoré meria údaje o hustote hmly, teplote a relatívnej vlhkosti. Tieto dôležité parametre sú potrebné pre statické a štatistické vyhodnotenie kvality prenosového prostredia pre FSO komunikáciu. Pre odhad veľkosti útlmu spôsobeného hmlou sa meria obsah skvapalnenej vody vo vzduchu v jednotkách g/m^3. Táto úloha si zvyčajne vyžaduje komplexné a drahé zariadenia. Pre experimentálne účely bol vyvinutý nízko nákladový prístroj, tzv. Senzor hmly (Obr. 1) [4].

Senzor pozostáva z vonkajšej a vnútornej jednotky. Vonkajšia jednotka meria požadované hodnoty, vnútorná jednotka konvertuje dáta z rozhrania RS 422 na rozhranie RS 232 a je napájaná zo siete napätím 12 V. Týmto napätím sa napája aj vonkajšia jednotka použitím

dátového kábla. Namerané údaje sú len bezrozmerné čísla, ktoré sa neskôr pri spracovaní pomocou vzorcov v programe prerátavajú na konkrétne veličiny charakterizujúce hmlu.



Obr. 1 Hardvérová schéma senzora hmly.

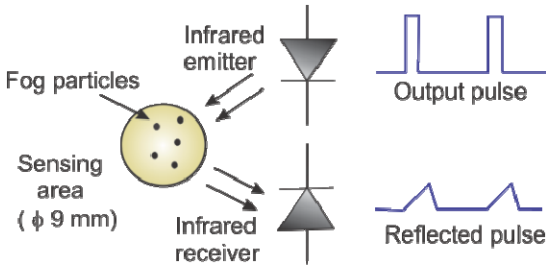
A. Princíp merania senzora hmly

Vonkajšia jednotka senzora hmly je zobrazená na Obr. 2. Je nainštalovaná na vonkajšej stene Laboratória Optoelektronických Systémov (LOS) na KEMT FEI TUKE.



Obr. 2 Vonkajšia jednotka senzora hmly.

Princíp merania hmly je zobrazený na Obr. 3. Vonkajšia jednotka pravidelne vysiela infračervené impulzy. Amplitúdy odrazených impulzov sú ovplyvnené nečistotami a jemnými čiastočkami v ovzduší. Práve to je použité pre výpočet obsahu skvapalnenej vody v ovzduší. Taktiež táto vonkajšia jednotka je vybavená senzormi teploty a relatívnej vlhkosti. Tieto tri parametre poskytujú dôležité informácie a majú charakteristické hodnoty pre hmlisté počasie. Senzor je uložený v skrinke, ktorá ho chráni pred priamym slnečným žiarením a dažďom, ale zároveň umožňuje voľný pohyb čiastočiek do a z oblasti senzora [4].



Obr. 3 Princíp merania hmlovej hustoty.

B. Technické vybavenie senzora hmly

Senzor hmly pozostáva z vonkajšej a vnútornej časti. Komunikácia medzi týmito dvoma časťami sa zabezpečuje cez port RS- 422 s nasledujúcimi parametrami:

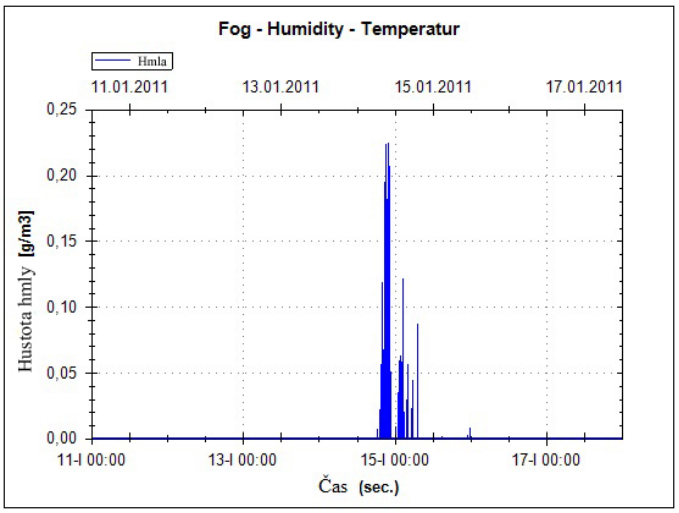
- 38400 kbit/s, s jedným stop bitom bez hardvérovej kontroly toku.
- Maximálna dĺžka kábla sa pohybuje okolo 1200 m [4].

Vonkajšia jednotka obsahuje mikropočítač, ktorý zberá merané dáta a komunikuje s vnútornou jednotkou. Firmware mikropočítača môže byť zmenený pomocou vnútornej jednotky a pomocou programu boot loader, ktorý komunikuje s počítačom pomocou sériového rozhrania. Boot loader môže byť použitý len s mikropočítačom, ktorý umožňuje zapisovať do svojej flash pamäte [4].

Dáta sú merané každých 10 ms, a sto nameraných údajov sa s priemerní čím dostaneme každú sekundu jednu hodnotu, ktorá sa zapíše do textového súboru. Okrem tejto hodnoty sú v textovom súbore zapísané aj hodnoty prislúchajúce pre teplotu, relatívnu vlhkosť, spriemerňujúcu hodnotu a relatívny čas. Spriemerňujúca hodnota má vplyv na počet vzoriek, ktoré sa zapisujú za celý deň do textového súboru. Ak by bola hodnota nastavená na 1000 a nie 100, tak by boli hodnoty zapísané len každých 10 sekúnd. Relatívny čas je vlastne poradie vzoriek v danom dni. Keďže program nám zapisuje hodnoty každú sekundu, tak relatívny čas je aj počet sekúnd od začiatku dňa.

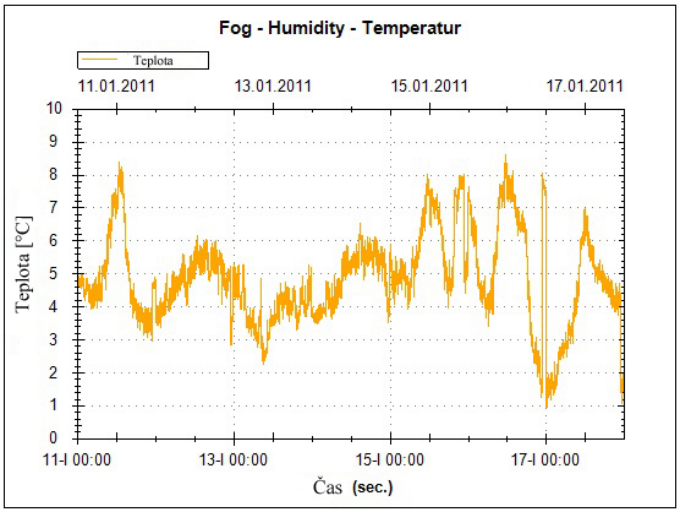
III. Experimenty so senzorom hmly

Namerané údaje je možné vykresliť pomocou programu na spracovanie nameraných údajov. Program umožňuje vykresliť jeden deň, niekoľko dní, dokonca aj niekoľko mesiacov. To je veľmi dôležité pre vyhodnocovanie dostupnosti FSO systémov z časového hľadiska. Na Obr. 4 je znázornené časové obdobie od 11. januára 2011 do 17. januára 2011. Tento graf nám poskytuje informácie o výskyte a hustote hmly v danom časovom úseku.

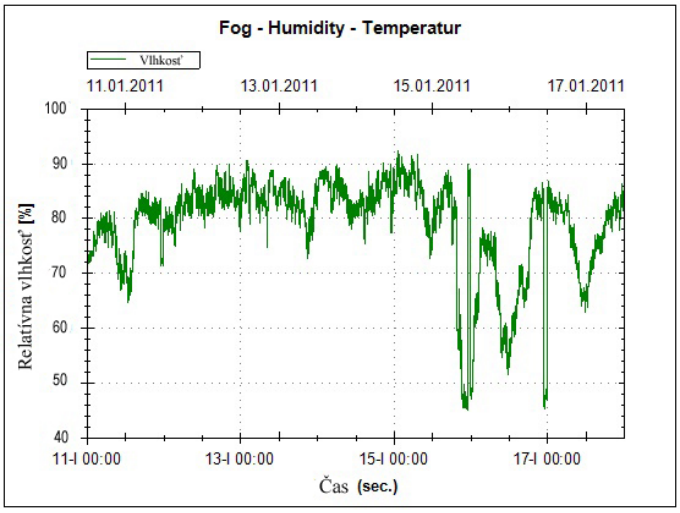


Obr. 4 Graf hustoty hmly.

Ako bolo spomenuté vyššie v kapitole II, je senzor vybavený aj snímačom pre meranie teploty a relatívnej vlhkosti prostredia. Výstupy nameraných hodnôt sú znázornené na Obr. 5 a Obr.6.



Obr. 5 Graf teploty prostredia.



Obr. 6 Graf relatívnej vlhkosti prostredia.

Podobne ako Obr. 4 aj Obr. 5 a Obr. 6 poskytujú dané informácie o parametroch kvality ovzdušia za časový interval od 11. do 17. januára 2011. Pri výbere údajov na vykreslenie sa používa kalendár, v ktorom sa zadáva počiatočný a posledný deň, ktorý sa má v programe spracovať. Žltá farba krivky predstavuje teplotu prostredia a po prepočte nameraných údajov sa teplota udáva v °C. Hodnoty zelenej krivky reprezentujúcej relatívnu vlhkosť vzduchu sú po prepočte údajov udávané v %.

IV. ZÁVER

V tomto článku je prezentovaný senzor hmly, ktorý meria a uchováva informácie o hustote hmly, teploty a relatívnej vlhkosti prostredia. Podrobne popisuje jeho vonkajšiu, vnútornú časť a technické vybavenie. Experimentálne merania sú získané z areálu Technickej Univerzity v Košiciach. Výstupné údaje z týchto meraní budú použité pre návrh optických prenosových systémov (FSO) v rámci areálu TUKE. Taktiež z týchto informácií je možné vypočítať pomocou Matlabu alebo C# dostupnosť a spoľahlivosť FSO systémov. Merania sa stále vykonávajú a sú priebežne spracovávané a uchovávané v našom laboratóriu (LOS KEMT FEI TUKE).

POĎAKOVANIE

Podporujeme výskumné aktivity na Slovensku/ Projekt je spolufinancovaný zo zdrojov EÚ. Tento článok bol vypracovaný v rámci projektu "Centrum excelentnosti integrovaného výskumu a využitia progresívnych materiálov a technológií v oblasti automobilovej elektroniky", ITMS 26220120055 (50%). Táto práca bola taktiež podporovaná a realizovaná z projektu Rozvoj Centra informačných a komunikačných technológií pre znalostné systémy (ITMS kód: 26220120030) na základe podpory operačného programu Výskum a vývoj financovaného z Európskeho fondu regionálneho rozvoja (50%).

LITERATÚRA

[1] O. Bouchet, et al. “Free Space Optics (Propagation and Communication),” ISTE Ltd., London, 2006.

[2] I. Kim, E. Korevar, “Availability of Free Space Optics (FSO) and Hybrid FSO/RF Systems,” Proc. SPIE, vol. 4530, 2001, 84-95.

[3] F. Nadeem, V. Kvicera, A. M. Saleem, E. Leitgeb, et al. “Weather Effects on Hybrid FSO/RF Communication Link,” IEEE Journal on Selected Areas in Communications, vol. 27, no. 9, December 2009, pp. 1687-1697.

[4] L. Csurgai-Horváth and J. Bitó, “Fog Attenuation on V Band Terrestrial Radio and a Low Cost Measurement Setup,” Future Network & Mobile Summit, Florence, Italy, Paper #47, June 2010.

Dostupnosť a spoľahlivosť FSO linky vzhľadom na viditeľnosť

Matúš TATARKO, Ľuboš OVSENÍK, Ján TURÁN

Katedra elektroniky a multimediálnych telekomunikácií, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

matus.tatarko@tuke.sk, lubos.ovsenik@tuke.sk, jan.turan@tuke.sk

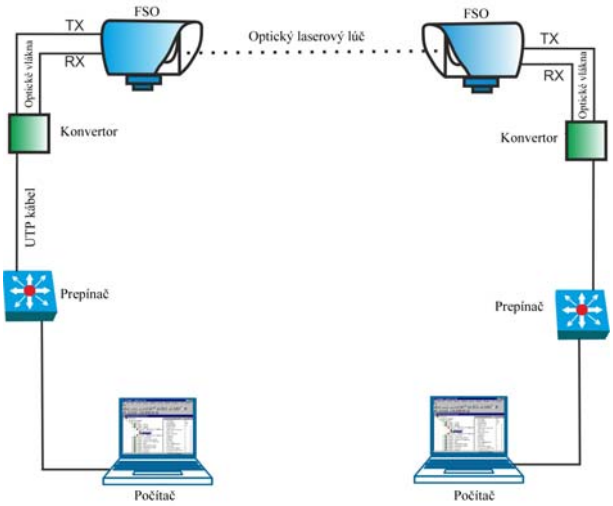
***Abstrakt* — Tento príspevok je zameraný na výpočet dostupnosti a spoľahlivosti FSO (Free Space Optics) systémov. FSO je optický systém, ktorý prenáša informácie voľným prostredím, teda vzduchom za pomoci zdroja optického signálu - lasera. Tento prenos informácií je umožnený medzi dvoma statickými bodmi, ktoré sú v priamej viditeľnosti. FSO systém môžeme tiež nazvať aj ako optický systém poslednej míle. Toto riešenie poslednej míle nepotrebuje žiadne drahé optické vlákna a vytvorenie takéhoto spojenia je veľmi jednoduché. Keďže sa jedná o systém, ktorého prenosovým médiom je vzduch, tak existuje množstvo atmosférických javov, ktoré majú vplyv na kvalitu služieb a dostupnosť tejto optickej linky. Medzi ne patrí rozptyl, absorpcia a turbulencie spôsobené veľkými výkyvmi prijatého výkonu a útlmom laserového lúča. Vplyvy absorpcie a turbulencie je možné korigovať správnym navrhnutím FSO linky. Napriek spomínaným negatívnym vplyvom, najväčší negatívny vplyv na kvalitu prenosového kanála má viditeľnosť. V typických vnútrozemských oblastiach, kde sa vyskytujú hmly, padajú rôzne zrážky, či už dážď alebo sneh je dôležité vedieť ich hodnoty. Tento článok popisuje zariadenie na meranie poveternostných podmienok a informácie o dostupnosti a spoľahlivosti FSO liniek na Slovensku.**

***Kľúčové slová* — dostupnosť, FSO systém, obsah skvapalnenej vody, viditeľnosť**

I. Úvod

Optická komunikácia voľným prostredím (FSO) je rozvíjajúca sa technológia, ktorá umožňuje plne duplexnú, protokolovo nezávislú komunikáciu medzi dvoma statickými bodmi. Pri týchto systémoch je potrebná priama viditeľnosť medzi vysielačom a prijímačom a vzdialenosť medzi nimi je od niekoľkých metrov až po niekoľko kilometrov. Umožňujú vysokorýchlostný prenos dát, až do niekoľko desiatok gigabajtov za sekundu. FSO systémy majú množstvo výhod, medzi ktoré patrí: rýchla a lacná inštalácia, nie sú potrebné drahé optické káble, žiadne plošné inštalácie a žiadne licencované spektrá. Dvojica FSO hlavíc môže prenášať veľké množstvo informácií bez vzájomných rušení medzi ostatnými FSO zariadeniami. Taktiež je ťažké odpočúvať takýto dátový prenos, pretože zväzok lúčov je malý a je možné ho aj kódovať.

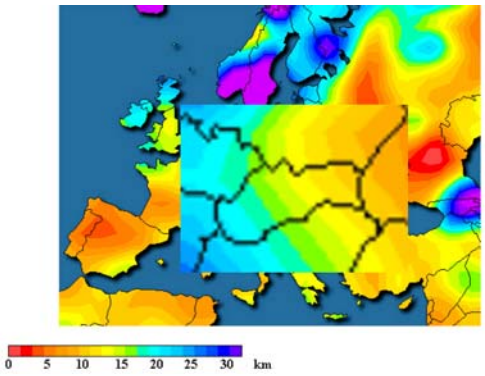
FSO systém pozostáva z dvoch hlavíc, pričom každá z hlavíc vysiela aj prijíma optický signál. Ak chceme zostaviť spojenie medzi hlavicami, t.j. aby nastala komunikácia medzi nimi, je potrebné ich vzájomné nasmerovanie a pripojenie k internetu, alebo k zdroju dát. Po úspešnom spojení, môžeme odpojiť jednu hlavicu od zdroja dát a spojenie medzi hlavicami sa už bude udržiavať. Na Obr.1 je znázornená hardvérová schéma FSO spojenia. Nevýhodou tohto systému je veľká zmena prijímaného výkonu spôsobená atmosférickými javmi. Tento fakt limituje dostupnosť a spoľahlivosť FSO pre daný prenosový rozsah. Takže je veľmi dôležité a dobré poznať poveternostné podmienky, medzi ktoré patria hmla, relatívna vlhkosť prostredia a teplota. Z týchto parametrov je možné následne vypočítať viditeľnosť, ktorá je daná v kilometroch. Laserové lúče týchto systémov sú pohlcované a rozptýlené vzdušnými molekulami a nečistotami, ktorých veľkosť je porovnateľná s vlnovou dĺžkou emitovaného optického žiarenia. Absorpcia spôsobuje zníženie intenzity signálu a rozptyl spôsobuje zmenu smeru optického lúča [1]. Tieto javy nastávajú počas hmlistých dní, kedy relatívna vlhkosť je vysoká a viditeľnosť nízka.



Obr. 1 Hardvérová schéma FSO systému.

II. INFORMÁCIE O VIDITEĽNOSTI

Informácie o viditeľnosti v Európe môžeme získať z internetových stránok o počasí, ktoré nám každodenne poskytujú informácie o hodnote viditeľnosti. Jedna z týchto internetovských stránok je aj http://www.wunderground.com/global/Region/EU/Visibility.html, ktorá nás informuje len graficky o hodnote viditeľnosti (Obr.2).



Obr. 2 Informatívna mapa z internetovej stránky.

Na Obr.2 môžeme vidieť len farebnú škálu s hodnotami, ktoré patria danej mape Európy. V prípade potreby presnej hodnoty viditeľnosti v danom mieste na Zemi nám takéto informácie nepomôžu. Taktiež všetky internetovské stránky s informáciami o počasí neposkytujú údaje o viditeľnosti, pretože zariadenia na meranie viditeľnosti sú drahé. Na Slovensku je len zopár takýchto meracích staníc, a to vo veľkých mestách alebo na letiskách.

Zariadenia pre priame meranie viditeľnosti sú drahé, existuje však aj iný spôsob na výpočet viditeľnosti. Je to výpočet pomocou parametra hmly, nazývaného obsah skvapalnenej vody vo vzduchu (LWC - Liquid Water Content). LWC popisuje množstvo vodných kvapôčiek v jednotkovom objeme hmly. Udáva sa v g/m^3. Priama metóda na meranie LWC pozostáva z extrahovania známeho objemu vzduchu cez bavlnenú podložku alebo pomocou pohára v rotujúcom zariadení a kondenzovaná voda sa následne odváži. Taktiež zmena odporu na horúcej drôtovej sonde pripevnenej k lietadlu pri prelete mrakmi vie poskytnúť hodnotu LWC. V hmle sa hodnota LWC pohybuje v širokom rozsahu. Vo väčšine modelov útlmu hmly je nutné poznať hodnotu LWC [5].

Inštalácia FSO systému vyžaduje znalosti prostredia, v ktorom bude budúci FSO systém pracovať. Na Technickej univerzite v Košiciach máme dva FSO systémy: prvý je od firmy FSONA s pracovnou vlnovou dĺžkou 1550 nm a druhý je od firmy Lightpointe, pracujúci na vlnovej dĺžke 850 nm. Z dôvodu rôznej vlnovej dĺžky má jeden hmlistý deň rôzne vplyvy na tieto linky. Z tohto dôvodu potrebujeme detailné informácie o počasí a hmle v našom areáli. Máme nainštalované zariadenie na meranie poveternostných parametrov (Senzor hmly) ako je hustota hmly, relatívna vlhkosť a teplota prostredia.

III. SENZOR HMLY

Tento senzor bol vytvorený ako jednoduché a nízko nákladové zariadenie na experimentálne použitie. Meria dôležité parametre pre ďalšie statické a štatistické vyhodnocovanie kvality voľného priestoru (vzduchu) pre optické komunikácie (FSO). Hmla má významný vplyv na kvalitu prenosu, pozostáva z vodnej pary alebo vodných čiastočiek s priemerom okolo 100 nm. Pre estimáciu veľkosti útlmu spôsobeného hmlou sa meria obsah skvapalnenej vody (LWC) v g/m^3.

Namerané údaje sú len bezrozmerné čísla, ktoré sa neskôr pri spracovaní pomocou vzorcov v programe prerátavajú na konkrétne jednotky a sú zobrazené v Tabuľke 1.

Tabuľka 1
Namerané údaje zo Senzora hmly

Hustota hmly	Teplota	Relatívna vlhkosť	Spriemerňujúca hodnota	Relatívny čas
227	4006	3121	100	0
217	4006	3121	100	1
213	4006	3121	100	2
213	4007	3212	100	3
211	4007	3121	100	4
216	4008	3121	100	5
216	4008	3121	100	6
258	4008	3122	100	7
216	4008	3122	100	8
218	4008	3122	100	9
218	4009	3122	100	10
215	4009	3122	100	11
215	4010	3122	100	12
215	4009	3119	100	13
210	4010	3119	100	14

Namerané údaje môžu byť následne spracované v počítači rôznymi spôsobmi. Môžu byť uložené ako textové súbory, ako tabuľky alebo uložené na serveri. V simulačnom prostredí Matlab vieme z hodnôt v prvom stĺpci tabuľky vypočítať hodnoty LWC, pre každú jednu nameranú vzorku.

A. *Spracovanie údajov v Matlabe*

Pre výpočet hustoty hmly používame údaje z prvého stĺpca Tabuľky 1:

$$F = (D_i \frac{5}{1024} - 0{,}96)\frac{0{,}5}{2{,}9} \tag{1}$$

kde F je hustota hmly a D_i sú výstupné hodnoty zo Senzora hmly. Ak chceme dostať presnú hodnotu skvapalnenej vody v ovzduší, potrebujeme definovať konštantu, ktorá nám pomôže prekonvertovať hodnotu hustoty hmly na hodnotu LWC. Táto konštanta je ďalej označená ako C a je to pomer medzi priemernými hodnotami obsahom skvapalnenej vody W_i a výstupnými hodnotami Senzora hmly D_i [2].

$$C = \frac{\sum_{i=1}^{n} W_i}{\sum_{i=1}^{n} D_i} = 0{,}7384 \quad \left[g / m^3\right] \tag{2}$$

Kde n je počet vzoriek počas celého výskytu hmly. Táto konštanta je použitá na zmenu momentálnych hodnôt snímačov D_i na momentálne hodnoty LWC [2]:

$$LWC = FC. \tag{3}$$

Z hodnôt LWC je už možné vypočítať viditeľnosť. Pre výpočet viditeľnosti je použitý základný vzťah pre viditeľnosť ako funkciu hustoty hmly:

$$V = d(LWC)^{-0{,}65} \tag{4}$$

kde V je viditeľnosť v kilometroch a parameter d nadobúda špecifické hodnoty podľa rôznych typov hmiel, ktoré sú uvedené v Tabuľke 2 [3,4]:

Tabuľka 2

Rôzne hodnoty parametra *d*

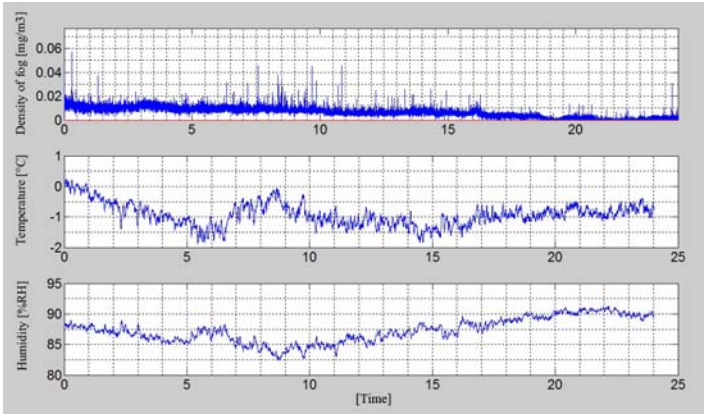
Typ hmly	Parameter *d*
Hustá hmly	0, 013
Vnútrozemská hmla	0, 034
Prímorská hmla	0, 060
Hustá selektívna hmla	0, 017
Stabilná a rozvíjajúca sa hmla	0, 024
Prenášajúca sa hmla	0, 02381

V našom prípade počítame s hodnotou parametra *d* pre stabilnú a rozvíjajúcu sa hmlu, teda rovnica má tvar:

$$V = 0{,}024(LWC)^{-0{,}65}. \tag{5}$$

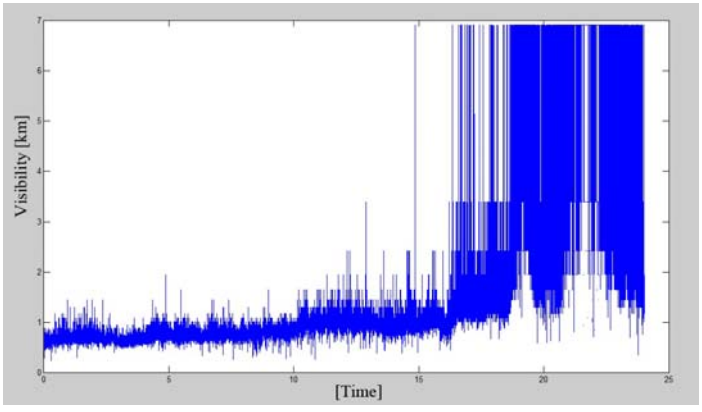
IV. EXPERIMENTÁLNE VÝSLEDKY

Po spracovaní zdrojových údajov v Matlabe je možné vykresliť dva druhy grafov. Prvý typ grafu je znázornený na Obr. 3 a obsahuje tri samostatné grafy o hustote hmly, teplote a relatívnej vlhkosti.



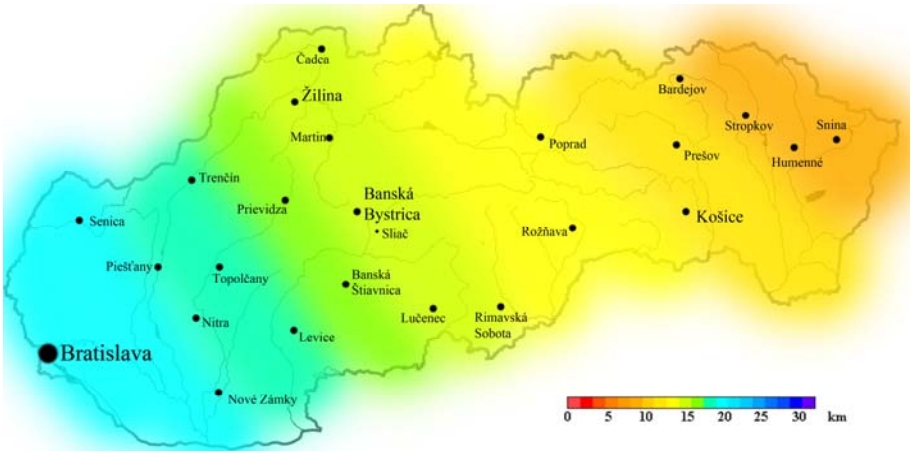
Obr. 3 Grafický výstup spracovaných údajov v Matlabe.

Druhým typom grafu je graf viditeľnosti, ktorý je znázornený na Obr. 4. Pre názornosť sú použité rovnaké zdrojové údaje ako pri Obr. 3.



Obr. 4 Viditeľnosť v Košiciach dňa 27. decembra 2011.

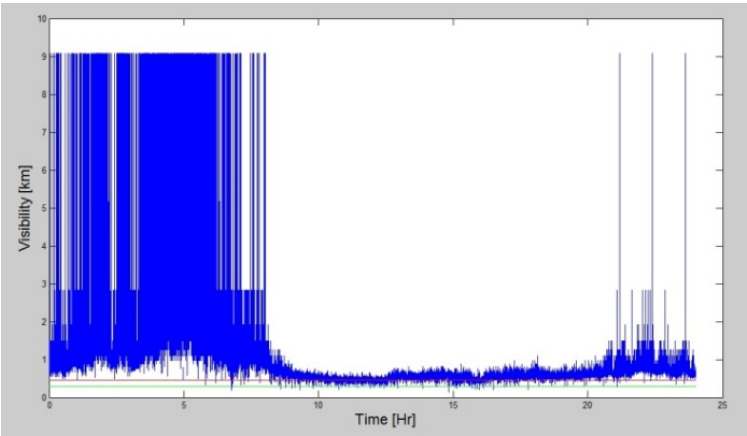
Na nasledujúcom Obr. 5 je farebná mapa Slovenska z 27. decembra 2011. Ide taktiež len o informatívnu mapu, ale aj napriek tomu môžeme vidieť, že viditeľnosť je v širokom rozsahu od 4 km do 20 km. To je dôvod, prečo je veľmi dôležité používať Senzor hmly na meranie parametrov ovzdušia. V každom mieste na mape sú rôzne podmienky pre prenos optického lúča.



Obr. 5 Farebná mapa Slovenska z 27. decembra 2011.

Získané informácie o viditeľnosti nám pomáhajú určiť dostupnosť a spoľahlivosť FSO liniek. V Matlabe môžeme zadať vzdialenosť alebo vzdialenosti medzi hlavicami vysielača a prijímača FSO systému. Záleží to od počtu simulovaných FSO liniek. V našom prípade simulujeme dve FSO linky (uvedené v kapitole II). Tieto FSO linky majú nasledovné vzdialenosti: linka FSONA má 451 m; linka Lightpointe má 312 m. Z týchto vzdialeností vyplýva, že aj pri minimálnej viditeľnosti budú tieto linky dostupné. Všetky hodnoty viditeľnosti, ktoré prekročia nami zadané hraničné hodnoty sa berú ako výpadok spojenia.

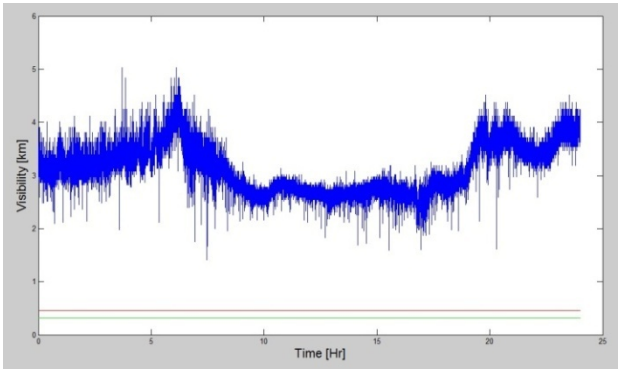
Z nameraných hodnôt zo Senzora hmly môžeme vypočítať aktuálnu viditeľnosť každú sekundu počas celého dňa. V Matlabe špecifikujeme hraničné hodnoty, t.j. vzdialenosti medzi hlavicami a graf viditeľnosti je vytvorený spolu so zadanými hranicami (Obr. 6). Pre každý systém program vytvorí textový dokument, v ktorom sú zaznamenané nižšie hodnoty ako nami zadané hraničné hodnoty.



Obr. 6 Graf viditeľnosti z 31. októbra 2011.

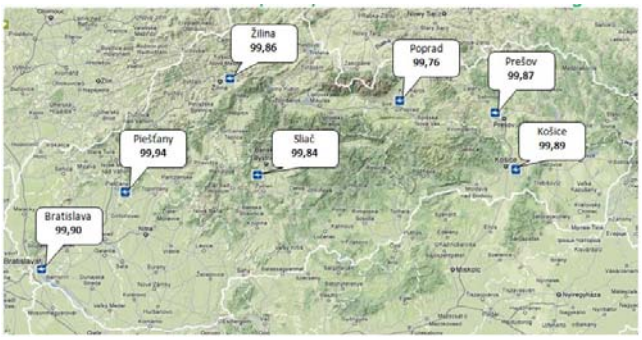
Na Obr. 6 sú znázornené aj dve hraničné čiary. Vrchná červená hraničná čiara patrí pre systém FSONA a spodná zelená hraničná čiara patrí pre systém Lightpointe. Možno povedať, že systém FSONA (SonaBeam 155E) má lepšiu dostupnosť. Vieme aj presne určiť, koľko percent z dňa, resp. nejakého časového intervalu linka padla. V tomto prípade vieme povedať, že linka vypadla na 3004 sekúnd, čo predstavuje 3, 471 % z celého dňa. Druhá linka vypadla len 90 sekúnd, čo predstavuje 0, 104 % z dňa. Meranie bolo získané počas hmlistého dňa 31. októbra 2011. Z toho vyplýva, že dostupnosť prvej FSO linky bola 96, 539 % a dostupnosť druhej FSO linky bola 99, 896 %.

Obr. 7 znázorňuje namerané údaje z 15. novembra 2011, kedy nebol zaznamenaný žiaden výskyt hmly. Všetky hodnoty viditeľnosti sú vyššie ako hraničné hodnoty oboch systémov. Z uvedeného vyplýva, že dostupnosť oboch liniek je 100 %.



Obr. 7 Graf viditeľnosti z 15. november 2011s hraničnými hodnotami.

Z nameraných údajov o viditeľnosti na slovenských letiskách je vytvorený Obr. 8, ktorý znázorňuje dostupnosť FSO liniek v % za roky 2010 a 2011.



Obr. 8 Priemerná dostupnosť FSO liniek na Slovensku.

V. ZÁVER

Článok nám dáva detailné informácie o viditeľnosti a metódach ako ju získať. Sú popísané poveternostné podmienky, ktoré negatívne vplývajú na FSO prenosový systém. Použitím senzora hmly a spracovaním nameraných údajov možno určiť dostupnosť FSO systémov na báze merania viditeľnosti. Tieto boli realizované z meraní v areáli TUKE, ako aj spracovaním dostupných údajov z vybraných letísk SR.

POĎAKOVANIE

Podporujeme výskumné aktivity na Slovensku/ Projekt je spolufinancovaný zo zdrojov EÚ. Tento článok bol vypracovaný v rámci projektu "Centrum excelentnosti integrovaného výskumu a využitia progresívnych materiálov a technológií v oblasti automobilovej elektroniky", ITMS 26220120055 (50%). Táto práca bola taktiež podporovaná a realizovaná z projektu Rozvoj Centra informačných a komunikačných technológií pre znalostné systémy (ITMS kód: 26220120030) na základe podpory operačného programu Výskum a vývoj financovaného z Európskeho fondu regionálneho rozvoja (50%).

LITERATÚRA

[1] I. I. Kim, E. Korevaar. "Availability of free space optics (FSO) and hybrid FSO/RF systems," Proc. of SPIE, vol. 4530, pp. 84-95, 2001.

[2] Z. Kolka, O. Wilfert, V. Biolkova. "Reliability of Digital FSO links in Europe," Int. J. Electronics, Communications, and Computer Engineering, vol. 1, no. 4, pp. 236–239, 2007.

[3] M. Reymann, J. Piasecki, F. Hosein and col. Meteorological Techniques, July 1998.

[4] S. Sheikh Muhammad, M. Saleem Awan, A. Rehman. Estimation and Liquid Water Content Based Attenuation Modeling for Fog in Terrestrial FSO Links. RADIOENGINEERING, Vol. 19, No. 2, June 2010.

[5] V. Zhurbenko. Electromagnetic Waves, June 2011, ISBN 978-953-307-304-0, p. 158-169.

Detekcia vodorovného značenia pomocou optického korelátora

Ján TURÁN, Ľuboš OVSENÍK, Tomáš HARASTHY

Katedra elektroniky a multimediálnych telekomunikácií, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

jan.turan@tuke.sk, lubos.ovsenik@tuke.sk, tomas.harasthy@tuke.sk

***Abstrakt* — Videodohľadové systémy sú v súčasnosti kvôli zvýšenému výskytu teroristických útokov stále aktuálnejšou témou. Sledovanie objektov vo videosekvenciách, resp. vo video streamingu je nevyhnutnou súčasťou videodohľadu. Pri dokonalom sledovaní objektu vo videu je potrebné rýchle spracovanie veľkého množstva informácie, ktorým disponujú videosekvencie. Práve tento problém je hlavným kameňom úrazu pri detekcii objektov v reálnom videu. Nulové , resp. minimálne oneskorenie medzi vstupným videom a videom s detegovaným objektom je nevyhnutné. Spracovanie video informácie s dostačujúcim rozlíšením v reálnom čase kladie vysoké nároky na klasické výpočtové metódy detekcie objektu vo videosekvenciách. Túto hlavnú nevýhodu systémov založených na elektronickej výpočtovej báze pomáhajú riešiť optické procesory ako napr. optický korelátor. Detekcia objektov má veľký význam aj v automobilovom priemysle, presnejšie pri videosistenčných systémoch vodiča. V tomto príspevku bude predstavený spôsob detekcie vodorovných dopravných čiar. Tu prezentovaný systém pomáha udržiavať automobil medzi čiarami a tak predísť nehodám spôsobených najmä nepozornosťou vodiča.**

***Kľúčové slová* — Cambridge korelátor, optická korelácia, spojené výkonové spektrum**

I. Úvod

Videodohľadové systémy vzbudzujú v súčasnosti veľký záujem. Rozpoznávanie objektov (či už tvári alebo iných biometrických prvkov) má obrovské využitie v mnohých odvetviach priemyslu aj bezpečnosti. Detekcia objektov v reálnom videu si vyžaduje spracovanie veľkého množstva informácie v reálnom čase. Za spracovanie v reálnom čase môžeme považovať také oneskorenie, ktoré nebude mať žiaden vplyv na plynulé spracovanie kontinuálneho videa [1,4].

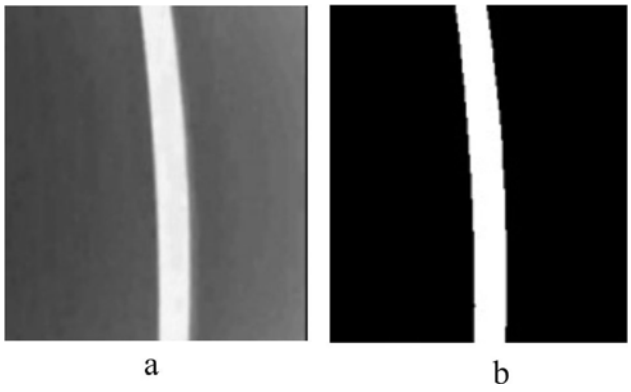
Optické korelátory dokážu spracovávať video informácie v reálnom čase a tak prispieť k vylepšeniu už navrhnutých systémov. V tomto príspevku bude popísaný optický korelátor, konkrétne Cambridge korelátor, ktorý sme využili v asistenčnom systéme vodiča na detekciu vodorovných dopravných čiar. Tento systém deteguje prítomnosť vodorovného značenia v blízkosti automobilu a pri vzdialení vozidla od stredovej čiary, resp. pri veľkom priblížení môže upozorniť vodiča.

V Kapitole 2 je popísaný spôsob predspracovania vstupného videa pomocou prahovania. V Kapitole 3 je v skratke popísaný Cambridge korelátor a princíp korelácie pomocou JPS korelátora. Experimenty s navrhnutým systémom spolu s výsledkami sú uvedené v Kapitole 4., zhrnutie a záver v Kapitole 5.

II. Prahovanie videa

Nosnou časťou pri realizácii tohto systému je predspracovanie videa pre detekciu vodorovného značenia. Toto predspracovanie sa realizuje softvérovo ešte pred samotnou koreláciou v Cambridge korelátore.

Zo vstupnej scény bolo vodorovné značenie zvýraznené pomocou techniky prahovania. Podstatou prahovania je upraviť každý obrazový prvok snímky buď do čiernej alebo bielej farby na základe voliteľnej úrovne jasu. Na Obr. 1vidíme vplyv prahovania obrazu snímaného spôsobom ktorý bude spolu s referenčným obrazom spracovávaný optickým korelátorom. V tomto prípade je kamera umiestnená pod spätným zrkadlom a snímanie je vedené kolmo na vozovku. Na Obr. 1a je uvedený príklad snímania obrazu týmto spôsobom (je snímaná plná čiara) a na Obr. 1b je jej prahovaná verzia. Pomocou vhodne zvolenej prahovej hranice dostaneme takýto obraz, ktorý je možné pomocou optického korelátora jednoducho spracovať [3].



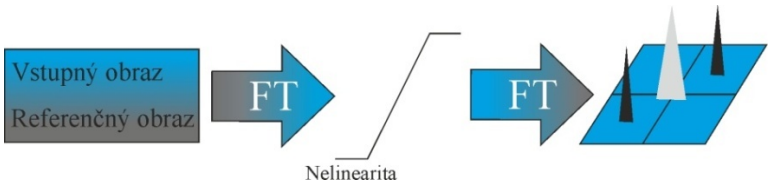
Obr. 1 Snímané vodorovné značenia, a) bez prahovania, b) prahované.

III. KORELÁCIA POMOCOU SPOJENEJ TRANSFORMÁCIE

Existujú dva základné druhy optických korelátorov [1]. Prvý typ je Korelátor s prispôsobeným filtrom (MFC). Jeho základ spočíva vo využití Vanderlugtovho filtra. Pri tomto druhu koralátorov sú dva obrazy, ktoré chceme porovnať Fourierovo transformované každý zvlášť. Dokonca sa Fourierova transformácia (FT) referenčného obrazu vykonáva mimo korelátora na vytvorenie referenčného filtra (odtiaľ aj názov prispôsobený filter). Tieto dva Fourierovo transformované obrazy sú opticky vynásobené a Fourierovo transformované ešte raz. Výsledkom druhej transformácie je korelácia dvoch vstupných obrazov. Zhoda týchto obrazov je reprezentovaná korelačnou špičkou intenzity jasu v korelačnej rovine.

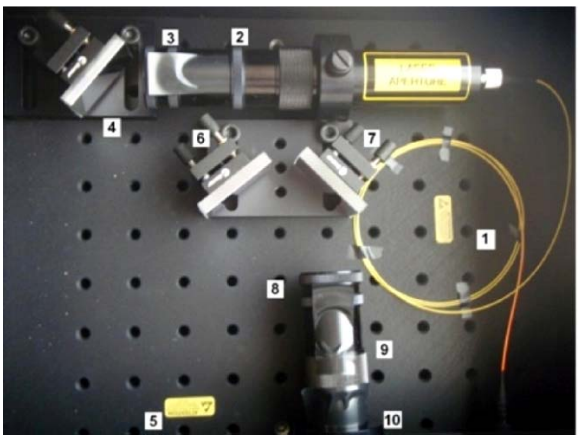
Ďalší typ využíva spojenú transformáciu, resp. spojené výkonové spektrum a je nazvaný Korelátor so spojenou transformáciou (JTC). Cambridge korelátor, ktorý je využitý v návrhu nášho systému patrí do tejto skupiny korelátorov [1,2].

Princíp korelácie s využitím spojenej transformácie je zobrazený na Obr. 2.



Obr. 2 Korelácia so spojenou transformáciou.

Pri korelátoroch so spojenou transformáciou sa vstupný a referenčný obraz netransformujú osobitne ale sú spojene do vstupnej scény a zobrazené na priestorovom modulátore svetla (SLM). Takto spojené obrazy sú FT spolu a vytvárajú Spojené výkonové spektrum (Joint Power Spectrum – JPS). Po úprave tohto spektra (väčšinou sa využíva prahovanie) je JPS znova Fourierovo transformované. Výsledkom druhej FT je autokorelácia vstupnej scény a teda korelačná rovina obsahuje autokorelačnú DC zložku a dvojicu korelačných špičiek, ktoré detegujú zhodu medzi vstupným a referenčným obrazom, ktoré boli spojené do vstupnej scény korelačného procesu [1,2].



1. Laserová dióda a jednovidové optické vlákno
2. Kolimačné šošovky
3. Polarizátor
4. Rovinné zrkadlo
5. Priestorový modulátor svetla
6. Rovinné zrkadlo
7. Rovinné zrkadlo
8. Analyzátor
9. Transformačné šošovky
10. CMOS senzor

Obr. 3 Korelácia so spojenou transformáciou.

Ako je spomenuté vyššie, Cambridge korelátor patrí do skupiny JTC korelátorov. Cambridge korelátor bol vyvinutý v spolupráci Cambridge University s firmou Cambridge Correlators Limited. Na Obr. 3 sú znázornené základné komponenty Cambridge korelátora. Jeho základom je tzv. „Fourier Transform Engine“, dômyselný návrh usporiadania šošoviek do tvaru písmena „W“ [2].

Priestorový modulátor svetla SLM (5) je uložený v prednej ohniskovej rovine systému šošoviek a je pripojený k počítaču (PC, laptop) prostredníctvom DVI rozhrania. Obraz na SLM je osvetlený laserovým lúčom z laserovej diódy. Svetlo je modulované obrazmi na SLM, a tak premietané prostredníctvom šošoviek. Tento systém šošoviek realizuje optickú Fourierovu transformáciu (OFT) v zadnej časti ohniskovej roviny a výstupná intenzita svetla je zachytená CMOS senzorom. Kolimačné a transformačné šošovky sú umiestnené v presne určených pozíciách, no pre jemnejšie nastavenie ich je možné doladiť. To isté platí aj pre rovinné zrkadlá, ktoré sa pomocou mikro-pohybu nastavovacích skrutiek dokážu dolaďovať pre dosiahnutie čo najlepších výsledkov. Otáčaním analyzátora a polarizátora okolo vlastnej osi je tiež možné nastaviť požadovanú fázu a amplitúdu optického výstupu [1,2,4].

IV. Experimenty s rozpoznávaním vodorovného značenia

Pomocou optického korelátora opísaného vyššie sme testovali navrhnutý systém na detekciu vodorovného dopravného značenia, konkrétne plnej čiary. Vodorovná čiara bola snímaná kamerou z automobilu. Kamera bola umiestnená tak aby čo najviac pripomínala kameru umiestnenú pod spätným zrkadlom. Vstupná snímka zosnímaná kamerou je zobrazená Obr. 1a a na Obr. 1b je jej prahovaná verzia.

Podľa takto nasnímaného videa sme tento vstupný obraz porovnávali s priamou čiarou. Vstupná scéna korelačného procesu teda pozostávala z nasnímanej snímky a referenčnej čiary. Takto vytvorená vstupná scéna bola vstupom pre Cambridge korelátor. Ako je spomenuté vyššie, výstupom korelácie pri korelátoroch so spojenou transformáciou pri detekcii zhody pozostáva z dvojice korelačných špičiek.

Na Obr. 4 je zobrazená výstupná korelačná rovina. Táto rovina obsahuje jednosmernú zložku umiestnenú v strede roviny. Táto jednosmerná zložka reprezentuje autokoreléciu celej vstupnej scény. Korelačná rovina obsahuje taktiež korelačné špičky, ktoré detegujú zhodu medzi referenčnou a vstupnou zložkou. V tomto prípade pre nás nie je zaujímavé to, že bola v korelačnom procese detegovaná zhoda, ale posun korelačných špičiek počas snímania videa.



Obr. 4 Korelácia so spojenou transformáciou.

Podľa toho ako sa čiara posúvala pri jazde autom, aj korelačné špičky vykazovali relatívne zhodný pohyb. Podľa tohto zistenia vieme určiť, že ak sa korelačné špičky posúvajú v nami vopred uvedených hraniciach, (červené štvorce) automobil zostáva medzi čiarami a teda nehrozí nebezpečenstvo čelnej zrážky s iným vozidlom (ak však druhé vozidlo nejazdí v protismere).

V tomto prípade bolo zvolených 10 oblastí kde sa korelačná špička môže nachádzať, resp. keď sa vozidlo nachádza medzi čiarami. Pre zjednodušenie by sme mohli zvoliť len jednu väčšiu oblasť a tak by sme určovali iba výskyt špičiek v kontrolovanej oblasti, t.j. ak sa špička v oblasti nachádza, vozidlo je medzi čiarami, ak nie ta je buď príliš v pravo alebo vľavo. Umiestnením viacerých oblastí vieme určiť ktorým smerom sa vozidlo pohybuje a tak nie len upozorniť vodiča na nebezpečenstvo ale napovedať mu do ktorej strany má vozidlo usmerniť. Pri tomto spôsobe by bolo tiež možné mierne vozidlo usmerniť aj bez zásahu vodiča.

V tejto aplikácii je optický korelátor využitý na detekciu čiary na vozovke a nie na porovnávanie rôznych obrazov. Na takéto využite optického korelátora sa hodia práve korelátory so spojenou transformáciu a to kvôli tomu, že majú menší odstup signálu od šumu (jednosmernej zložky od korelačných špičiek). Pri zvyšovaní počtu objektov vo vstupnej rovine

sa zvyšuje aj jednosmerná zložka a korelačné špičky sú ťažšie detekovateľné. Naopak využitie korelátora s prispôsobeným filtrom by v tomto prípade nebolo vhodné. Korelátor s prispôsobeným filtrom vyžaduje vytvorenie prispôsobeného filtra pomocou optickej Fourierovej transformácie. Vytvorenie OFT pomocou elektroniky by spôsobilo vyššie oneskorenie a teda by nemusel systém upozorniť vodiča včas, alebo by bolo možné vytvoriť OFT pomocou ďalšieho optického korelátora, čo je nepraktické.

V. ZÁVER

V tomto príspevku sme prezentovali využitie optického korelátora vo videoasistenčnom systéme vodiča na detekciu vodorovného značenia. V experimentoch sme využili Cambridge korelátor.

V úvode príspevku je v skratke popísane potrebné predspracovanie vstupnej scény ktoré je nevyhnutné pre redukciu nežiaducich zložiek obrazu. Keďže sme sa zamerali výlučne na detekciu plnej vodorovnej čiary, ako predspracovanie nám poslúžilo len binárne prahovanie. Takto predspracovanú vstupnú snímku sme pomocou optického korelátora porovnávali s referenčnou snímkou.

V ďalšej kapitole bol v stručnosti predstavený Cambridge korelátor, ktorý sme využili v našom experimente. Cambridge korelátor patrí do skupiny korelátorov so spojenou transformáciou (JTC) a tie naproti korelátorom s prispôsobeným filtrom nepotrebujú žiadne informácie vopred na vytvorenie prispôsobeného filtra. Vďaka tejto vlastnosti Camridge korelátora a všeobecne korelátorov so spojenou transformáciou je vhodnejšie využívať tento typ korelátorov pri detekcii objektov v reálnom čase.

Ďalšou výhodou korelátorov so spojenou transformáciou je, že dokážu vykonávať viacero optických Fourierových transformácií naraz, bez oneskorenia. Počet transformácií je obmedzený iba rozlíšením priestorového modulátora svetla (SLM) a rozlišovacou schopnosťou kamery, ktorá sníma výstup SLM. Táto výhoda JTC korelátorov umožňuje detegovať aj prerušované aj plné čiary naraz, resp. systém doplniť o detekciu rôznych šípok a výstražných vodorovných značení.

Nasledujúci výskum so spracovaním videa pomocou Cambridge korelátora bude smerovať práve týmto smerom. Využite viacnásobnej optické korelácie bez zvýšenia oneskorenia má veľký význam a umožňoval by tak spojiť systém detekcie vodorovného značenia so systémom na rozpoznávanie dopravných značiek. Spojenie týchto systémov by si v tomto prípade žiadalo použite druhej kamery alebo zmeniť spôsob snímania vodorovného značenia použitím vhodného optického systému tak, aby bolo možné pomocou jednej kamery snímať zvislé aj vodorovné značenie.

POĎAKOVANIE

Podporujeme výskumné aktivity na Slovensku/ Projekt je spolufinancovaný zo zdrojov EÚ. Tento článok bol vypracovaný v rámci projektu "Centrum excelentnosti integrovaného výskumu a využitia progresívnych materiálov a technológií v oblasti automobilovej elektroniky", ITMS 26220120055 (50%). Táto práca bola taktiež podporovaná a realizovaná z projektu Rozvoj Centra informačných a komunikačných technológií pre znalostné systémy (ITMS kód: 26220120030) na základe podpory operačného programu Výskum a vývoj financovaného z Európskeho fondu regionálneho rozvoja (50%).

LITERATÚRA

[1] J. W. Goodman, *Introduction to Fourier Optics*. Colorado: Roberts & Company ISBN 0-9747077-2-4, 2005.
[2] Cambridge Correlators,: “http://www.cambridgecorrelators.com,“ April 2012.
[3] S. E. Embauch, *Computer Vision and Image Processing a practical approach using CVIPtools.* Prentice Hall PTR, Upper Saddle River, NJ 07458, June 1999.
[4] F. T. S. Yu – X. Yang,: *Introduction to Optical Engineeringgnition.* Cambridge University PressISBN 0-521-57493-5 1997.

Fepstrálna analýza v systéme automatického rozpoznávania reči

Peter VISZLAY, Miroslav MUCHANIČ, Jozef JUHÁR

Katedra elektroniky a multimediálnych telekomunikáci, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

peter.viszlay@tuke.sk, miroslav.muchanic@student.tuke.sk, jozef.juhar@tuke.sk

***Abstrakt* — Tento článok je zameraný na opis a aplikáciu pomerne nového typu analýzy rečového signálu - fepstrálnej analýzy v systémoch automatického rozpoznávania reči (ARR). Úvod článku poskytuje krátky prehľad súčasného stavu v oblasti aplikácie metód analýzy reči v systémoch ARR. Ďalšie kapitoly obsahujú podrobnejší opis fepstrálnej analýzy a jej aplikácie. Záver článku prezentuje experimentálne výsledky úspešnosti aplikácie fepstrálnej analýzy v úlohe rozpoznávania plynulej reči pre anglický jazyk s malým slovníkom na báze foném. Experimenty sú porovnané s referenčnými modelmi založenými na konvenčných Mel-frekvenčných kepstrálnych koeficientoch (MFCC).**

***Kľúčové slová* — extrakcia príznakov, fepstrálna analýza, Mel-frekvenčné kepstrálne koeficienty, systém ARR**

I. Úvod

Posledné desaťročia bola vynaložená veľká snaha vedcov na to, aby dokázal človek plnohodnotne komunikovať s počítačom hovorenou rečou. Ani za 50 rokov sa však nepodarilo docieliť to, aby systém vedel rozpoznať akéhokoľvek rečníka a v rôznorodom prostredí. Pre spoľahlivú komunikáciu človeka so strojom je potrebné vykonať viacero krokov. Jedným z nich je aj predspracovanie rečového signálu, kedy sa signál upraví na postupnosť príznakových vektorov. Rečový signál prechádza procesom digitalizácie a pre získanie vhodnej reprezentácie sa využívajú tzv. metódy krátkodobej analýzy. Rečový signál sa kontinuálne spracováva po krátkych časových úsekoch – mikrosegmentoch (obvykle 30*ms*), ktoré sú extrahované aplikáciou vybraného typu oknovej funkcie. V ďalšom kroku sa prevádza spektrálna analýza, ktorá je často sprevádzaná aplikáciou banky vhodných filtrov [1]. Filtre sú navrhnuté takým spôsobom, aby modelovali nelineárne vlastnosti ľudského sluchu.

II. Extrakcia príznakov v Systémoch Automatického Rozpoznávania Reči

Medzi konvenčné metódy extrakcie príznakov reči patrí LPCC metóda (Linear Prediction Cepstral Coefficients) využívajúca modelovanie frekvenčného spektra vstupného signálu pomocou lineárne predikovaných koeficientov. Ďalšou je percepčna lineárna predikcia (PLP) založená na modeli ľudského sluchového vnímania. Najrozšírenejšou je MFCC analýza (Mel-frekvenčná kepstrálna analýza), ktorej koeficienty Melovského kepstra sa vypočítajú pomocou diskrétnej kosínusovej transformácie (DCT) pre dané výkonové pásma Melovského spektra [1].

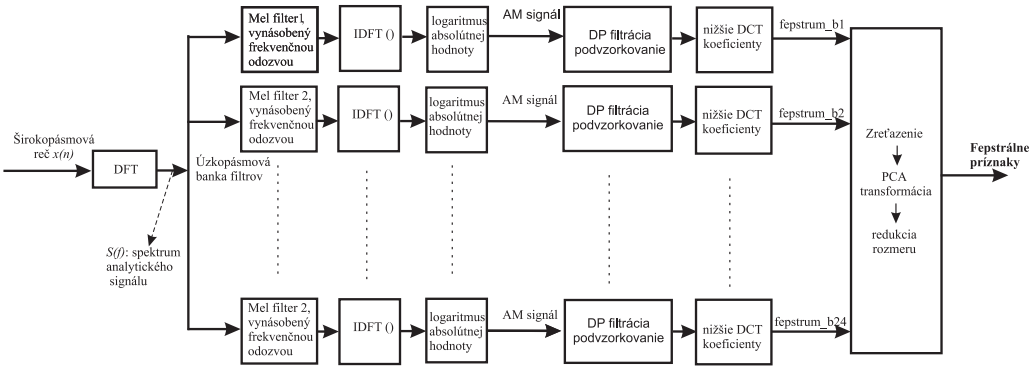
Medzi nové metódy extrakcie príznakov je možné zaradiť metódu PMVDR (Perceptual Minimum Variance Distortionless Response). Táto metóda je založená na procedúre, ktorá odhaduje prenosové charakteristiky zo spektra signálu vypočítaním tzv. vyššej spektrálnej obálky nazývanej MVDR spektrum [2][3]. Metóda PNCC (Power Normalized Cepstral Coefficients) koeficientov je veľmi podobná metóde MFFC, s tým rozdielom, že MFCC transformácia je nahradená tzv. Gammatone filtrom, ktorý simuluje správanie sa slimáka v ľudskom uchu [4][5]. V [6] autori opisujú metódu TRAPS (TempoRAl Patterns), kde obvyklý spektrálny vektor príznakov je nahradený 1*s* dlhým časovým vektorom logaritmických spektrálnych energií v kritickom pásme [7].

Ďalšou, pomerne novou metódou (navrhnutá v roku 2006) je fepstrálna analýza [2], ktorá vykonáva amplitúdovú moduláciu na veľmi úzkej banke filtrov. Doposiaľ bola testovaná len na anglickom jazyku. Fepstrálna analýza je predmetom aj tejto publikácie.

III. Fepstrálna analýza Rečového Signálu

Fepstrálna analýza je metóda odhadu parametrov založená na algoritme, ktorý vykonáva AM dekompozíciu rečového signálu v časovej oblasti. Má veľa spoločných čŕt s kepstrálnou metódou spomínanou v predchádzajúcej časti. Fepstrum je v podstate iným pomenovaním pre reálne kepstrum. Demodulácia v časovej oblasti vedie k robustnému odhadu plynulého, pomaly sa vyvíjajúceho AM signálu, Pre odhad ktorého je dôležité zabezpečiť, aby bol signál úzkopásmový. Bloková schéma fepstrálnej analýzy je znázornená na Obr. 1.

Z tejto schémy vyplýva, že proces fepstrálnej analýzy je možné rozdeliť na dve časti, a to pred a po rozložení signálu na subpásma. V prvej fáze, kým nie je signál rozložený, sa po konverzii rečového signálu na postupnosť diskrétnych vzoriek, vykoná niekoľko operácií. Po načítaní signálu sa ako prvá vykoná segmentácia na rámce. Na vytvorenie rámca sa používa pravouhlá oknová funkcia. Dĺžka okna je pri fepstrálnej analýze 100ms. Po narámcovaní signálu, tento pokračuje ďalej do bloku, v ktorom sa vykoná diskrétna Fourierova transformácia (DFT). Keďže spektrum signálu po DFT transformácii je symetrické, vezme sa do úvahy iba pozitívne frekvenčné spektrum, čím sa získa polovičné množstvo vzoriek ako pri pôvodnom rečovom signáli.



Obr. 1 Bloková schéma fepstrálnej analýzy

V ďalšej časti procesu analýzy je širokopásmový signál rozložený do 24 Mel-frekvenčných subpásiem. Rozdelenie na subpásma je vykonané filtráciou Melovskou bankou filtrov. Filtrácia sa vykonáva jednoduchým prenásobením frekvenčnej odozvy každého trojuholníkového filtra so vzorkami po DFT analytického signálu *S(f)*. Potom použitím inverznej DFT transformácie sa získa filtrovaný analytický signál pre každé pásmo. Výsledkom aplikácie logaritmu na absolútnu hodnotu pre každý signál v každom pásme je amplitúdovo modulovaný signál. Tento AM signál je následne filtrovaný cez 40 bodový filter s pohyblivým priemerom (tzv. moving average filter) a následne je signál podvzorkovaný faktorom 40. Následne sa vykoná diskrétna kosínusová transformácia (DCT), z ktorej sa zachová len prvých 5 koeficientov s nižším rádom. Zreťazením vektorov z každého pásma vznikne fepstrálny supervektor s rozmerom 120, ktorý reprezentuje jeden rámec reči. Takáto fepstrálna reprezentácia je potom vstupom do PCA transformácie. Z tohto 120-rozmerného supervektora po vykonaní transformácie sa ponechá len polovica z celkového počtu parametrov. Tieto reprezentujú už finálny kompozitný fepstrálny príznakový vektor. Dominantných 60 vlastných hodnôt predstavuje takmer 95% z celkovej váhy všetkých vlastných hodnôt a zároveň predstavuje aj konečný počet parametrov, ktoré sú výsledkom fepstrálnej analýzy [2][8][9][10][11].

IV. Experimenty a Výsledky

A. Podmienky experimentov

Za účelom evaluácie metódy bola zvolená anglická akustická databáza TIMIT [12]. Trénovací korpus obsahuje približne 10 hodín spojitej reči nahranej od rôznych rečníkov a testovacia množina obsahuje približne 2 hodiny rečových dát. Ako evaluačný parameter bol zvolený parameter presnosti vyjadrený v percentách, ktorý je vypočítaný ako pomer správne rozpoznaných slov ku všetkým referenčným slovám.

Rečové signály pre referenčný model boli parametrizované podľa konvenčnej MFCC analýzy (preemfáza, Hammingovo okno na 30*ms* rámci, 10*ms* krok oknovania, FFT, Mel filtrácia, DCT, výpočet delta a akceleračných koeficientoch).

Pre porovnanie výsledkov boli použité referenčné MFCC modely. Tieto modely boli trénované a testované na základe 39-rozmerných MFCC vektorov ako ľavo-pravé skryté

Markovove modely, ktorých stavy boli modelované Gaussovskými zmesami až do počtu 256. Extrakcia príznakov pre referenčné modely a akustické modelovanie a testovanie boli vykonané s využitím nástroja HTK (Hidden Markov Model Toolkit). Fepstrálna analýzy bola implementovaná v prostredí MATLAB.

B. Akustické modely založené na fepstre

Bolo vykonaných niekoľko experimentov založených na fepstre, ktoré sa líšili vo veľkosti a type okna. Výsledkom pre každý experiment bola sada desiatich modelov, ktoré sa navzájom odlišovali v dimenzii fepstrálnych príznakov. Išlo o modely s nasledujúcim počtom dimenzií: *dim_120* (úplná PCA transformačná matica), *dim_90*, *dim_60*, *dim_39*, *dim_20*, *dim_13* a modely *dim_20_d_a*, *dim_20_d*, *dim_13_d_a*, *dim_13_d* (vznikli rozšírením základných modelov o delta a akceleračné koeficienty).

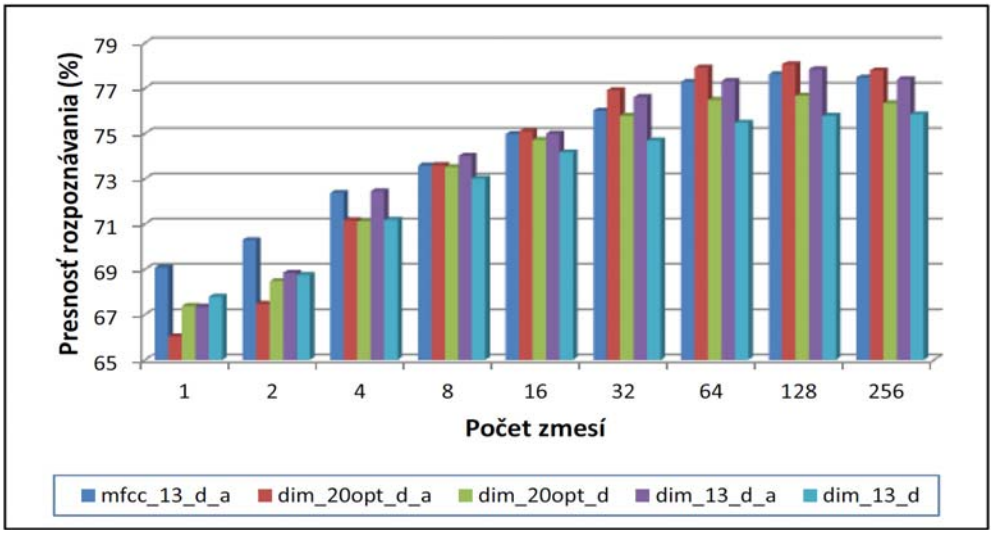
C. Výsledky experimentov

V tejto kapitole sú prezentované výsledky testovania fepstrálnej analýzy. V Tabuľke 1 sú uvedené hodnoty presnosti rozpoznania pre akustický model s Hammingovým oknom, dĺžkou rámcov 30*ms* a prekrývaním 20*ms*. Táto konfigurácia bola podľa výsledkov najvhodnejšia, výsledky ostatných modelov preto nie sú uvedené.

Tabuľka 1
Porovanie presnosti rozpoznania referenčného a fepstrálnych modelov

Počet zmesí	Referenčný model mfcc_13_d_a	Fepstrálne modely dim_20_d_a	dim_20_d	dim_13_d_a	dim_13_d
1	69,08	66,04	67,38	67,34	67,79
2	70,29	67,48	68,48	68,84	68,75
4	72,38	71,17	71,12	72,45	71,18
8	73,59	73,61	73,52	74,01	73,02
16	74,98	75,10	74,71	75,00	74,17
32	76,01	76,91	75,79	76,61	74,69
64	77,29	77,92	76,48	77,32	75,48
128	77,62	78,06	76,67	77,84	75,78
256	77,47	77,79	76,34	77,40	75,85

Na Obr. 2 sú znázornené uvedené hodnoty v grafe, kde tmavomodrou farbou je znázornený referenčný model. Zlepšenie je viditeľné hlavne v prípadoch, ktoré obsahujú delta aj akceleračné koeficienty (na grafe znázornené červenou a fialovou farbou).

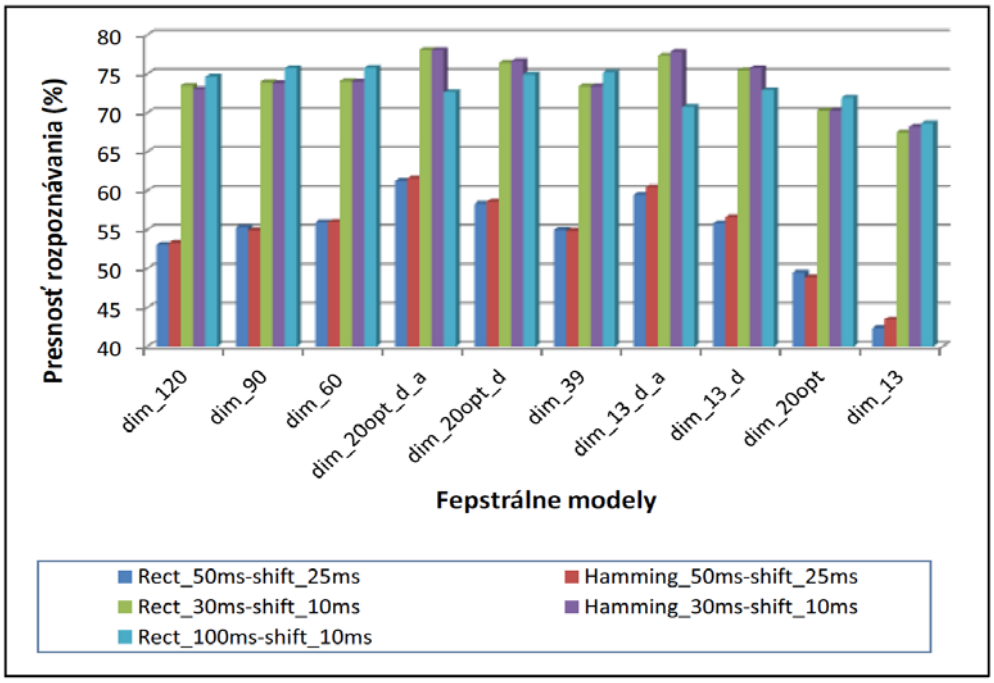


Obr. 2 Grafické znázornenie presnosti rozpoznania referenčného a fepstrálnych modelov

Na Obr. 3 je znázornená presnosť rozpoznania pre jednotlivé fepstrálne modely pre všetky vykonané experimenty pri použitých 128 Gaussových zmesiach. Ako je možné vidieť z grafu, najmenej uspokojivé výsledky boli dosiahnuté pri použití 50*ms* dlhých rámcov a to pri všetkých fepstrálnych modeloch. Model *Rect_100ms-shift_10ms* (model s použitým pravouhlým oknom a 100*ms* dlhými rámcami) dosahoval lepšie výsledky v prípadoch, kedy boli použité delta a akceleračné koeficienty. Pri modeloch rozšírených o tieto koeficienty mali lepšie výsledky modely s 30*ms* dlhými oknami a to pri použití ako pravouhlého, tak aj Hammingovho. Tieto experimenty v súhrne dosahovali najlepšie výsledky.

V. ZÁVER

Tento článok sa zameriava na fepstrálnu analýzu rečového signálu. Z vykonaných experimentov pre rôzne parametre (veľkosť okna, typ okna, prekrývanie rámcov, výstupná dimenzia vektorov) bola zistená presnosť rozpoznania slov v rozmedzí 20-78%. Hodnoty presnosti rozpoznania slov pre navrhnuté modely boli porovnané s hodnotami referenčného modelu. Vo väčšine prípadov boli zistené hodnoty nižšie v porovnaní s týmto referenčným modelom, avšak pri experimente so vstupnými parametrami 30*ms* okno s 15*ms* prekrývaním bolo dosiahnuté absolútne zlepšenie ~1% pri Hammingovom aj pravouhlom okne.



Obr. 3 Porovnanie jednotlivých fepstrálnych modelov pre parameter 128 zmesí

POĎAKOVANIE

Táto práca bola vytvorená realizáciou projektu Rozvoj Centra informačných a komunikačných technológií pre znalostné systémy (kód ITMS projektu: 26220120030) na základe podpory operačného programu Výskum a vývoj financovaného z Európskeho fondu regionálneho rozvoja.

LITERATÚRA

[1] J. Psutka, J. Matoušek, L. Müller, V. Radová, *Mluvíme s počítačem česky*. Praha: Academia, 2006, s. 156-167.
[2] S. Dharanipragada, U. H. Yapanel, B. D. Rao, „Robust Feature Extraction for Continuous Speech Recognition Using the MVDR Spectrum Estimation Method," IEEE Transaction on Audio, Speech, and Language Processing, Vol. 15, No. 1, January 2007, 11 s.
[3] G. Sárosi, M. Mozsáry, P. Mihajlik, T. Fegyó, „Comparison of Feature Extraction Methods for Speech Recognition in Noise-Free and in Traffic Noise Environment," 2011 6th Conference on Speech Technologoy and Human-Computer Dialogue (SpeD), s. 1-8.
[4] V. Tyagi, *Novel speech processing techniques for robust automatic speech recognitions: PhD Thesis.* Lausanne: EPFL, 2006, s. 100.
[5] Ch. Kim, R. M. Stern, „Feature extraction for robust speech recognition using a power-law nonlinearity and power-bias subtraction," IEEE Internation Conference on Acoustics Speech and Signal Processing (ICASSP), 14-19 March 2010, s. 28-31.
[6] H. Hermansky, S. Sharma, „Temporal patterns (TRAPS) in ASR of nosiy speech," Oregon Graduate Institute of Science and Technology, Portland, Oregon. USA. 1999, s. 289-292.
[7] F. Grézl, *Trap-based probabilistic features for automatic speech recognition.* Disertační práce. Brno: Vysoké učení technické v Brně. 2007, s.13-14.
[8] V. Tyagi, „An improved modulation spectrum for ASR." In the Proc. of Interspeech 2007. Antwerp, Belgium.
[9] V. Tyagi, Ch. Wellekens „Fepstrum and carrier signal decomposition of speech signals through homomorphic filtering," 2006 IEEE International Conference on Acoustics, Speech and Signal Processing, 2006. 14-19 May 2006.
[10] V. Tyagi, Chr. Wellekens, „Fepstrum Representation of Speech Signal," 2005 IEEE Workshop on Automatic Speech Recogntion and Understanding, 27 November 2005, 6 s.
[11] V. Tyagi, „Fepstrum Features: Design and Application to Conversational Speech Recognition," Computer Science, June 2011, 13 s.
[12] J. Garofolo, J. Lamel, W. Fisher, J. Fiscus, D. Pallett, N. Dahlgren, „The DARPA TIMIT acoustic-phonetic continuous speech corpus cd-rom", Technical report, 1993.

Banky filtrov pre implementáciu diskrétnej multiwaveletovej transformácie

Jozef ZAVACKÝ, Ján MIHALÍK

Katedra elektroniky a multimediálnych telekomunikácií, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

jozef.zavacky@tuke.sk, jan.mihalik@tuke.sk

***Abstrakt* — V konvenčných waveletových transformáciach nie je potrebná prefiltrácia vďaka dolnopriepustnej vlastnosti mierkovej funkcie. Toto však neplatí pre multiwaveletové transformácie. Niekoľko vedeckých článkov pre návrh prefiltrov už bolo opublikovaných, ale existujúce prefiltre sú obvykle neortogonálne, čo často spôsobuje problém pri kódovaní. Okrem toho podmienky kladené na prefiltre boli kladené len na prvý krok priamej diskrénej multiwaveletovej transformácie. V tomto článku navrhujeme prefilter, ktorý kombinuje myšlienky konvenčnej waveletovej transformácie a multiwaveletovej transformácie. Prefiltre sú ortogonálne, avšak nie sú maximálne decimované. Sú odvodené pre diskrétnu multiwaveletovú transformáciu navrhnutú Donovanom, Geronimom, Hardingom a Massopustom (DGHM).**

***Kľúčové slová* — multiwavelety, mnohonásobné rozlíšenie, mierkové a waveletové funkcie, rozkladové matice, multiwaveletová banka filtrov, ortogonalita, prefiltrácia, rád prefiltra, kombinované filtre**

I. ÚVOD

Za uplynulé dve desaťročia sa waveletové transformácie stali čoraz viac a viac populárnymi. Sú užitočným nástrojom pre také aplikácie spracovania jedno a dvojrozmerných signálov ako je kompresia, rozpoznávanie, analýza, odšumenie atď [1-6]. Je to vďaka skutočnosti, že wavelety umožňujú získať v rôznych časoch a pri rôznych frekvenciách odlišné rozlíšenie. Multiwavelety možno považovať za zovšeobecnenie skalárnych waveletov. Skalárne wavelety majú jednu mierkovú funkciu $\varphi(t)$ a waveletovú funkciu $\psi(t)$, zatiaľ čo multiwavelety majú dve, alebo viac mierkových a waveletových funkcií. Všeobecne diskrétna multiwaveletová transformácia (DMWT) môže mať *r* mierkových a *r* waveletových funkcií, vtedy hovoríme, že má násobnosť (multiplicitu) *r*. Väčšina DMWT používa dve mierkové a waveletove funkcie, pričom *r* môže mať teoreticky ľubovoľnú hodnotu [7]. Multiwaveletový rozklad môže byť realizovaný bankami filtrov, pričom rozkladové koeficienty v tomto prípade sú matice namiesto skalárov. Také vlastnosti ako je ortogonalita, symetria (t.j. lineárna fázová frekvenčná chrakteristika), vysoký rád aproximácie a krátky nosič (support), resp. časový interval na ktorom trvajú sú základom pre ich úspešné aplikovanie pri spracovaní signálov. Súčasné splnenie uvedených vlastností, ale nie je možné dosiahnúť pre skalárne wavelety.

II. DISKRÉTNA MULTIWAVELETOVÁ TRANSFORMÁCIA

Podobne ako v prípade skalárnej diskrétnej waveletovej transformácie, teória diskrétnej multiwaveletovej transformácie je založená na príncípe analýzy signálov s mnohonásobným rozlíšením. Mierkové a waveletové funkcie môžu byť zapísané ako stĺpcové vektory, t.j. multimierková funkcia $\boldsymbol{\varphi}(t) = \left[\varphi_1(t)\, \varphi_2(t) \ldots \varphi_\ell(t) \ldots \varphi_r(t)\right]^T$ a rovnakým spôsobom môžeme definovať multiwaveletovú funkciu $\boldsymbol{\psi}(t) = \left[\psi_1(t)\, \psi_2(t) \ldots \psi_\ell(t) \ldots \psi_r(t)\right]^T$, kde T znamená transpozíciu. Analogicky ako skalárna mierková a waveletová funkcia aj multimierková $\boldsymbol{\varphi}(t)$ a multiwaveletová $\boldsymbol{\psi}(t)$ funkcia musia vyhovovať nasledujúcim maticovým (dilatačným) rovniciam [8-9]

$$\boldsymbol{\varphi}(t) = 2\sum_k \tilde{\mathbf{g}}_0(k)\,\boldsymbol{\varphi}(2t-k) \tag{1}$$

$$\boldsymbol{\psi}(t) = 2\sum_k \tilde{\mathbf{g}}_1(k)\,\boldsymbol{\varphi}(2t-k), \quad k \in Z \tag{2}$$

kde namiesto skalárnych rozkladových koeficientov sú $\tilde{\mathbf{g}}_0(k)$ a $\tilde{\mathbf{g}}_1(k)$ rozkladové matice rozmeru $r \times r$. Potom dvojica týchto matíc $\tilde{\mathbf{g}}_0(k), \tilde{\mathbf{g}}_1(k)$ popisuje multiwaveletovú banku filtrov, pričom $\mathbf{g}_0(k) = \sqrt{2}\,\tilde{\mathbf{g}}_0(-k)$ je impulzná charakteristika vektorového dolnopriepustného (DP) filtra a $\mathbf{g}_1(k) = \sqrt{2}\,\tilde{\mathbf{g}}_1(-k)$ vektorového hornopriepustného (HP) filtra. Potom ich komplexné frekvenčné charakteristiky sú matice

$$\mathbf{G}_0(\mathrm{e}^{\mathrm{j}\omega}) = \sum_k \mathbf{g}_0(k)\,e^{-j\omega k} \quad \text{a} \quad \mathbf{G}_1(\mathrm{e}^{\mathrm{j}\omega}) = \sum_k \mathbf{g}_1(k)\,e^{-j\omega k} \tag{3}$$

pričom ich prvky sú trigonometrické polynómy. O multiwaveletovej banke filtrov (MWBF), s konečnou impulznou odozvou (KIO) hovoríme vtedy, ak existuje take celé číslo N, že $\mathbf{g}_0(k) = 0$ a $\mathbf{g}_1(k) = \mathbf{g}_1(k) = 0$ pre $k > N$. Označme stĺpcové vektory mierkových a detailových koeficientov DMWT $\mathbf{c}_j(k) = \left[(c_{j,1}(k) \ \ldots \ c_{j,r}(k)\right]^{\mathrm{T}}$, $\mathbf{d}_j(k) = \left[(d_{j,1}(k) \ \ldots \ d_{j,r}(k)\right]^{\mathrm{T}}$.

Potom aproximačné (mierkové) koeficienty pre jednu úroveň priamej DMWT sa vypočítajú takto

$$\mathbf{c}_j(k) = \sqrt{2}\sum_n \tilde{\mathbf{g}}_0(n-2k)\,\mathbf{c}_{j+1}(n) \tag{4}$$

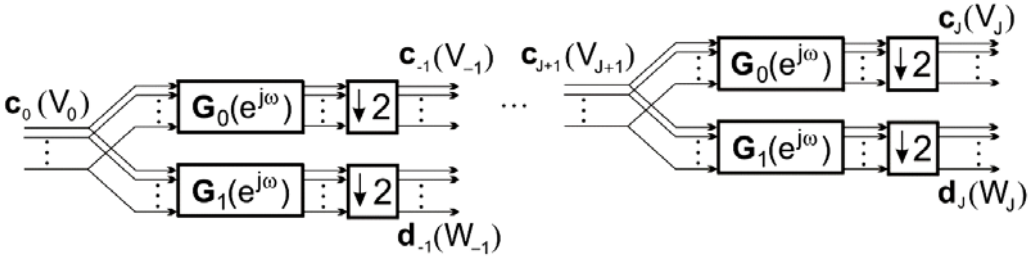
a detailové koeficienty

$$\mathbf{d}_j(k) = \sqrt{2}\sum_n \tilde{\mathbf{g}}_1(n-2k)\,\mathbf{c}_{j+1}(n) \ . \tag{5}$$

Výpočet spätnej DMWT pre jednu úroveň dostaneme je vyjadrený nasledovne

$$\mathbf{c}_{j+1}(n) = \sqrt{2}\sum_k \tilde{\mathbf{g}}_0^T(k-2n)\,\mathbf{c}_j(k) + \sqrt{2}\sum_k \tilde{\mathbf{g}}_1^T(k-2n)\,\mathbf{d}_j(k) \ . \tag{6}$$

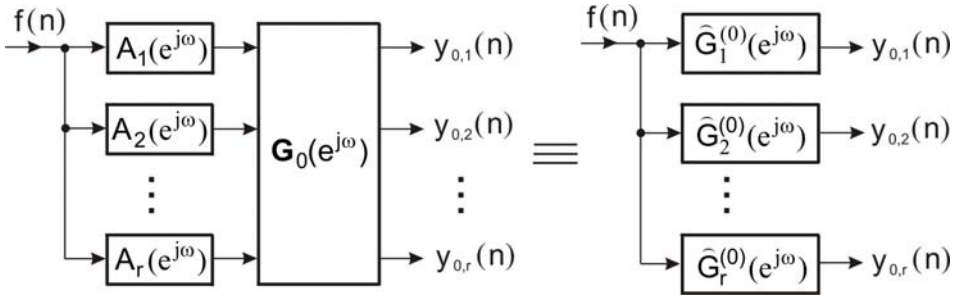
Pre výpočet $\mathbf{c}_j(k)$ a $\mathbf{d}_j(k)$, $J \le j \le -1$ potrebujeme určiť iba koeficienty $\mathbf{c}_0(n)$ zo vzoriek vstupného diskrétneho signálu *f(n)*. Implementácia úplnej priamej DMWT je na Obr. 1.



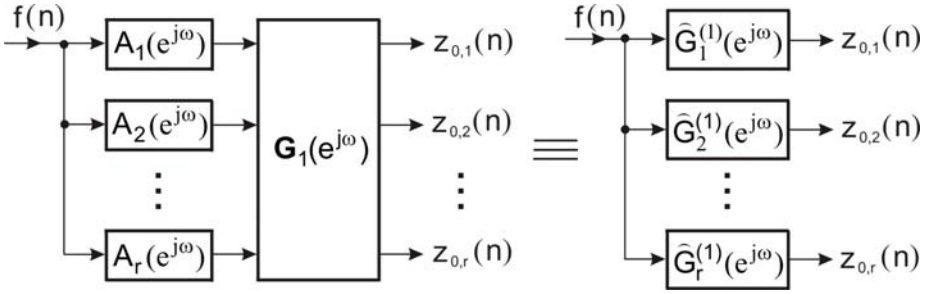
Obr. 1 Implementácia úplnej priamej DMWT.

III. Banky Filtrov pre Diskrétnu Multiwaveletovú Transformáciu

Jednou z dôležitých úloh pri aplikovaní DMWT je úloha jej inicializácie (prefiltrácie), keď je potrebné na začiatku obvykle existujúci iba jeden vstupný skalárny (jednorozmerný) signál f(n) pretransformovať vhodným spôsobom na r-rozmerný vstupný vektorový signál, ktorým je postupnosť vektorov začiatočných koeficientov $\mathbf{c}_0(n) = \left[(c_{0,1}(n) \ \ldots \ c_{0,r}(n)\right]^T$ [10]. Operácia pomocou ktorej sa táto transformácia vykoná sa nazýva prefiltrácia a realizovaná je prefiltrami. Príslušnú postfiltráciu je potom potrebné vykonať po spätnej DMWT a musí to byť inverzný proces k prefiltrácii. Pre frekvenčne nevyvážené vektorové filtre DMWT existuje niekoľko metód návrhu prefilrov [11]. Vhodnou prefiltráciou sa v tomto prípade kompenzuje nedostačujúca vlastnosť zachovania/ potlačenia signálov, resp. mierkových a detailových koeficientov. Druh použitej prefiltrácie je rozhodujúci pre získanie dobrých výsledkov



a)



b)

Obr. 2 Kombinované filtre a) $\widehat{G}_{\ell}^{(0)}(e^{j\omega})$ z prefiltra $A_{\ell}(e^{j\omega})$ a vektorového filtra $\mathbf{G}_0(\mathrm{e}^{\mathrm{j}\omega})$, b) $\widehat{G}_{\ell}^{(1)}(e^{j\omega})$ z prefiltra $A_{\ell}(e^{j\omega})$ a vektorového filtra $\mathbf{G}_1(\mathrm{e}^{\mathrm{j}\omega})$.

pre určitú aplikáciu DMWT. Komplexné frekvenčné charakteristiky $A_{\ell}(e^{j\omega})$, $\ell = 1,2,...,r$ prefiltrov typu KIO na Obr. 2. môžu byť rozložené (faktorizované) nasledovným spôsobom [12]

$$\begin{bmatrix} A_1(e^{j\omega}) \\ . \\ . \\ . \\ A_r(e^{j\omega}) \end{bmatrix} = \mathbf{U}_{\mu}(\omega) \cdots \mathbf{U}_1(\omega) \begin{bmatrix} A_1(1) \\ . \\ . \\ . \\ A_r(1) \end{bmatrix} \tag{7}$$

a

$$\mathbf{U}_s(\omega) = \mathbf{I}_r + (e^{-j\omega} - 1)\mathbf{u}_s^{\dagger}\mathbf{u}_s \tag{8}$$

kde $\mathbf{u}_s = (u_{s1},...,u_{sr})$ a norma vektora $\mathbf{u}_s$ je 1, t.j.,

$$\sum_{\ell=1}^{r} |u_{s\ell}|^2 = 1 \ . \tag{9}$$

Jediným obmedzením pre parametre $u_{s\ell}$ je, že musia spĺňať podmienku (9) pre $s = 1,2,...,\mu$. Parameter μ určuje dĺžku prefiltra a nazýva sa rád prefiltra s frekvenčnými charakteristikami $A_{\ell}(e^{j\omega})$, $\ell = 1,2,...,r$. Ak neexistuje žiadna frekvenčná charakteristika $\mathbf{U}_s(\omega)$ v rov.(8), potom $\mu = 0$, t.j., rád prefiltra je nulový. Pre DGHM DMWT (r = 2) s prefiltráciou prvého rádu rov.(7) nadobudne tvar

$$\begin{bmatrix} A_1(e^{j\omega}) \\ A_2(e^{j\omega}) \end{bmatrix} = \left[I_2 + (e^{-j\omega} - 1) \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix} \begin{bmatrix} \cos\theta & \sin\theta \end{bmatrix} \right] \times \begin{bmatrix} \sqrt{2} \\ 1 \end{bmatrix} \frac{1}{\sqrt{3}} \tag{10}$$

kde $\theta = \frac{1}{2} arctg \frac{\sqrt{2}}{2}$. Potom kombinované filtre budú vyjadrené následovne

$$\widehat{G}_{\ell}^{(0)}(e^{j\omega}) = \sum_{m=1}^{r} G_{\ell,m}^{(0)}(e^{j\omega})\, A_m(e^{j\omega}), \quad \ell=1,2,\ldots,r \tag{11}$$

.

$$\widehat{G}_{\ell}^{(1)}(e^{j\omega}) = \sum_{m=1}^{r} G_{\ell,m}^{(1)}(e^{j\omega})\, A_m(e^{j\omega}), \quad \ell=1,2,\ldots,r \tag{12}$$

kde $\mathbf{G}_0(e^{j\omega}) = [G_{\ell,m}^{(0)}(e^{j\omega})]_{r\times r}$ a $\mathbf{G}_1(e^{j\omega}) = [G_{\ell,m}^{(1)}(e^{j\omega})]_{r\times r}$.

IV. Experimentálne Výsledky pre Dghm Dmwt

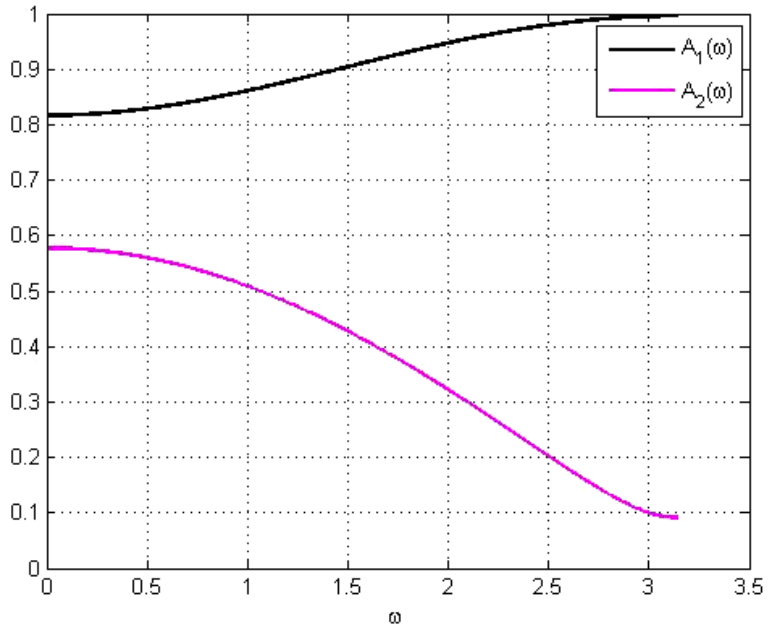
Jednou z najznámejších DMWT s násobnosťou $r = 2$ je DMWT navrhnutá Donovanom, Geronimom, Hardingom a Massopustom (DGHM) [13]. Mierkové a waveletové funkcie (bázové funkcie) tejto DMWT poskytujú kombináciu ortogonality, symetrie a kompaktného suportu (krátkeho nosiča), čo nemôže byť dosiahnuté pomocou žiadnej skalárnej waveletovej bázy. V tomto prípade $\boldsymbol{\varphi}(t) = [\varphi_1(t)\ \varphi_2(t)]^T$ a $\boldsymbol{\psi}(t) = [\psi_1(t)\ \psi_2(t)]^T$. Kde maticové impulzné charakteristiky $\mathbf{g}_0(\mathrm{k})$ a $\mathbf{g}_1(\mathrm{k})$ sú matice rozmeru 2×2, t.j. [11]

$$\mathbf{g}_0(0) = \begin{bmatrix} \frac{3}{10} & \frac{2\sqrt{2}}{5} \\ -\frac{\sqrt{2}}{40} & -\frac{3}{20} \end{bmatrix}, \quad \mathbf{g}_0(1) = \begin{bmatrix} \frac{3}{10} & 0 \\ \frac{9\sqrt{2}}{40} & \frac{1}{2} \end{bmatrix}, \quad \mathbf{g}_0(2) = \begin{bmatrix} 0 & 0 \\ \frac{9\sqrt{2}}{40} & -\frac{3}{20} \end{bmatrix},$$

$$\mathbf{g}_0(3) = \begin{bmatrix} 0 & 0 \\ -\frac{\sqrt{2}}{40} & 0 \end{bmatrix} \quad \mathbf{g}_1(0) = \begin{bmatrix} -\frac{\sqrt{2}}{40} & -\frac{3}{20} \\ -\frac{1}{20} & -\frac{3\sqrt{2}}{20} \end{bmatrix}, \quad \mathbf{g}_1(1) = \begin{bmatrix} \frac{9\sqrt{2}}{40} & -\frac{1}{2} \\ \frac{9}{20} & 0 \end{bmatrix}$$

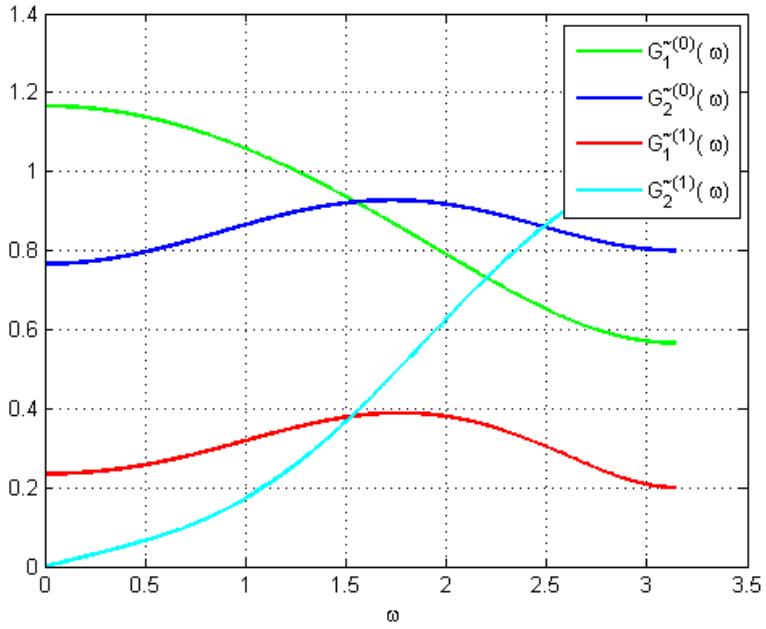
$$\mathbf{g}_1(2) = \begin{bmatrix} \frac{9\sqrt{2}}{40} & -\frac{3}{20} \\ -\frac{9}{20} & \frac{3\sqrt{2}}{20} \end{bmatrix}, \quad \mathbf{g}_1(3) = \begin{bmatrix} -\frac{\sqrt{2}}{40} & 0 \\ \frac{1}{20} & 0 \end{bmatrix}. \tag{13}$$

Na Obr. 3 sú zobrazené amplitúdové frekvenčné charakteristiky banky prefiltrov pre DGHM DMWT



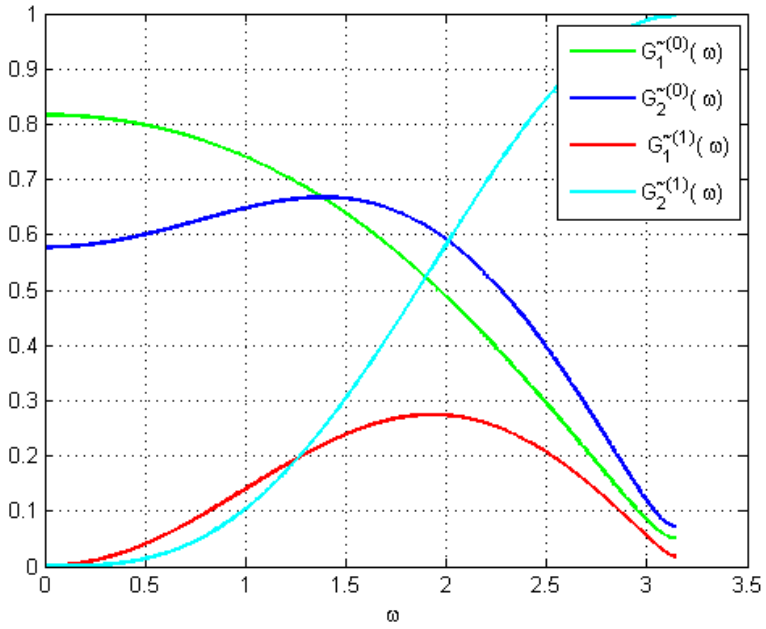
Obr. 3 Amplitúdové frekvenčné charakteristiky prefiltrov $A_1(\omega)$, $A_2(\omega)$ pre DGHM DMWT.

Na Obr. 4 sú zobrazené amplitúdové frekvenčné charakteristiky kombinovaných filtrov pre DGHM DMWT bez prefiltrácie $\widetilde{G}_1^{(0)}, \widetilde{G}_2^{(0)}, \widetilde{G}_1^{(1)}, \widetilde{G}_2^{(1)}$.



Obr. 4 Amplitúdové frekvenčné charakteristiky kombinovaných filtrov pre DGHM DMWT bez prefiltrácie $\widetilde{G}_1^{(0)}, \widetilde{G}_2^{(0)}\ \widetilde{G}_1^{(1)}, \widetilde{G}_2^{(1)}$.

Na Obr. 5 sú zobrazené amplitúdové frekvenčné charakteristiky kombinovaných filtrov $\widetilde{G}_1^{(0)}(e^{j\omega})$, $\widetilde{G}_2^{(0)}(e^{j\omega}), \widetilde{G}_1^{(1)}(\omega), \widetilde{G}_2^{(1)}(\omega)$ pre DGHM DMWT s prefiltráciou prvého rádu.



Obr. 5 Amplitúdové frekvenčné charakteristiky kombinovaných filtrov pre DGHM DMWT s prefiltráciou prvého rádu $\widetilde{G}_1^{(0)}, \widetilde{G}_2^{(0)}, \widetilde{G}_1^{(1)}, \widetilde{G}_2^{(1)}$.

Z porovnania priebehov amlitúdových frekvenčných charakteristík kobinovaných prefiltrov na Obr. 4 a 5 je zrejmé, že kobinované prefiltre na Obr. 5 majú lepšie vlastnosti typu DP a HP ako kombinované prefiltre na Obr. 4.

V. ZÁVER

V tomto článku sme navrhli banku prefiltrov pre DGHM DMWT s prefiltráciou prvého rádu. Následne sme vypočítali kombinované filtre z prefiltrov $A_1(e^{j\omega})$, $A_2(e^{j\omega})$ a vektorových filtrov $\mathbf{G}_0(e^{j\omega})$, $\mathbf{G}_1(e^{j\omega})$. Metóda návrhu prefiltrov je založená na aproximácii mierkovej funkcie s vlastnosťou dolnej priepuste a ortogonality jej posunutých verzií použitím lineárnych kombinácií multimierkových funkcií a ich translácií. Navrnutá banka prefiltrov je ortogonálna, ale nie je maximálne decimovaná. Je aplikovaná na všetky kroky rozkladu pre diskrétnu multiwaveletovú transformáciu, zatiaľ čo prefiltrácia napr. v [10] je sústredená iba na prvý krok priamej DMWT.

LITERATÚRA

[1] M. Raabbani - R. Joshi, An Overview of the JEPEG 2000 Still Image Compression Standard. Signal Processing Image Communication, 2002, 17, pp. 3-48.
[2] J. Zavacký - J. Mihalík - I. Gladišová, Implementácia diskrétnej waveletovej transformácie v štandarde JPEG-2000. Slaboproudý obzor,roč. 63, č. 3-4, 2007, s. 5-9, ISSN 0037-668X.
[3] S. Grace Chang - Yu. Bin - M. Vattereli, Adaptive Wavelet Threshold for Image Denoising and Compression. IEEE Trans., Image Processing, Vol. 9 Sept. 2000, pp. 1532-46.
[4] K. G. Kyriakopoulos - D. J. Parish, Applying wavelets for the controlled compression of communication network measurements. IET Communications, Vol. 4, No. 5, 2010, pp. 507-520.
[5] J. Bonal,- S. Donohue - C. McNally, Wavelet analysis of bender element signa. Geotechnique, Vol. 62, No. 3, 2012, pp. 243-252.
[6] M. F. Fahmy - G. Fahmy - O. F. Fahmy, B-spline wavelets for signal denoising and image compression. Signal Image and video processing, Vol. 5, No. 2, 2011, pp. 141-153.
[7] F. Keinert, Wavelets and Multiwavelets. Champan &Hall/CRC, 2004, ISBN 1-58-488-304-9.
[8] J. Y. Tham - L. Shen, - S. L. Lee - H. H. Tan, A General Approach for Analysis and Application of Discret Multiwavelet Transforms. IEEE Trans. On Signal Processing, Vol. 48, No. 2, 2000, pp. 457-464.
[9] J. Mihalík - J. Zavacký, Diskrétne spracovanie signálov, LČSOV FEI TU Košice, 2011, ISBN 978-80-553-0730-5, (319 strán).
[10] X. G. Xia - J. S. Geronimo - D. P. Harding - B. W. Suter, Design of Prefilters for Discrete Multiwavelet Transforms. IEEE Trans. On Signal Processing, Vol. 44, No. 1, 1996, pp. 25-34.
[11] S. Hongli - C. Yuanli - Q. Zulian, On Design of Multiwavelet Prefilters. Applied Mathematics and Computation, Vol. 172, 2006, pp. 1175-1187.
[12] T. C. Hsung - M. C. Sun - D. P. K. Lun - W. C. Siu, Symmetric prefilters for multiwavelets. IEE Proc.-Vis. Image Signal Process., Vol. 150, No. 1, 2003, pp. 59-68.
[13] G. Donovan - J. S. Geronimo - D. P. Harding - P. R. Massopust, Construction of Orthogonal Wavelets Using Fraction Interpolation Functions. SIAM J. Math. Anal., Vol. 27, 1996, pp. 1158-1192.

Celočíselná diskrétna waveletová transformácia

Jozef ZAVACKÝ

Katedra elektroniky a multimediálnych telekomunikácií, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

jozef.zavacky@tuke.sk

***Abstrakt* — Článok sa zaoberá implementáciou diskrétnej waveletovej transformácie jednorozmerných diskrétnych signálov. Navrhuje lifting metódu implementácie celočíselnej diskrétnej waveletovej transformácie (CČDWT) s použitím biortogonálnych filtrov s nepárnymi dlžkami impulzných charakteristík, ktoré sa líšia o nepárny násobok dvoch. Vykonaná bola imlementácia priamej CČDWT pre banku biortogonálnych filtrov (5,3) na jednotlivých stupňoch pre zvolené vstupné signály a výsledky sú zbrazené dvomi spôsobmi.**

***Kľúčové slová* — lifting metóda, predikcia, korekcia, aproximačne a detailové koeficienty, celočíselná diskrétna waveletová transformácia**

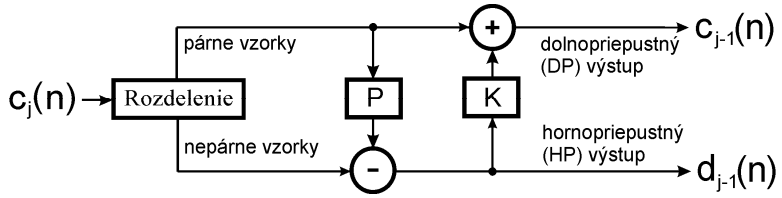
I. Úvod

Problematika waveletovej transformácie a jej aplikácií sa v posledných rokoch teší značnej pozornosti. Svedčí o tom rada publikácií zameraných na túto problematiku. Môžeme sa stretnúť s monotematicky zameranými publikáciami čiste matematickými [1], [2], s knihami písanými z užšieho pohľadu číslicového spracovania signálov [3], [4], [5], alebo s publikáciami v ktorých sa zmiešavajú obidva pohľady [6], [7]. V tomto článku uvádzame lifting metódu [8], [9] implementácie DWT, ktorá používa biortogonálne filtre (5,3). Táto metóda v porovnaní s klasickou konvolučnou metódou [3], [10] má nasledujúce výhody. Urýchluje výpočet, preto je tiež označovaná ako rýchla lifting DWT. Dáva možnosť vykonať výpočty bez použitia prídavnej pamäte, to znamená, že pôvodný diskrétny signál môže byť priamo nahradený jeho DWT. Spätná DWT môže být získaná bezprostredne z priamej DWT, zámenou poradia operácií a výmenou každého znamienka plus za mínus a naopak. Umožňuje ľahko získať celočíselnú DWT, ktorá má jednoduchú hardverovú implementáciu a umožňuje bezstratovú kompresiu. V mnohých aplikáciach číslicového spracovania signálov vstupné údaje pozostávajú z celočíselných vzoriek. Pritom waveletové koeficienty majú hodnoty s pohyblivou rádovou čiarkou, aj keď vstupné vzorky sú celé čísla, pretože koeficienty filtrov v zapojení procesorov DWT sú racionálne alebo celé čísla [11]. Celočíselné údaje sú veľmi žiadúce pre rýchlu implementáciu DWT, najmä v hardvérovej realizácii, kvôli ich účinnému zapamätaniu a kódovaniu.

II. Celočíselná Diskrétna Waveletová Transformácia

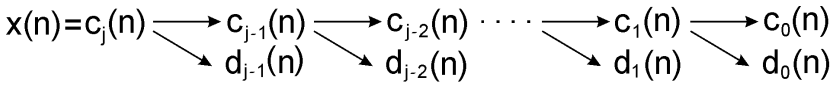
Lifting metóda (LM) implementácie DWT začína s triviálnym waveletom a postupne zostavuje nový wavelet so zlepšenými vlastnosťami pridávaním nových bázovych funkcií. To bolo zdrojom inšpirácie pre názov „lifting metóda“, čo by sme mohli nazvať metóda postupného vylepšovania vlastností nového waveletu. Uvažujme diskrétny signál $c_j(n)$ s 2^j vzorkami, ktorý chceme transformovať na jeho hrubšiu aproximáciu $c_{j-1}(n)$ a na detail $d_{j-1}(n)$. Typický prípad DWT realizovanej pomocou LM pozostáva z troch krokov : rozdelenie, predikcia a korekcia. Tieto tri kroky, ktoré súčasne zodpovedajú (5,3) DWT sú zobrazené v blokovej schéme na Obr.1, kde P je operátor predikcie a K operátor korekcie. V bloku rozdelenia sa vstupný signál $c_j(n)$ rozdelí na párne vzorky a nepárne vzorky. Vo všeobecnosti rozdelenie na podmnožiny párnych a nepárnych vzoriek nie je jediné možné, ale pokiaľ chceme využiť autokoreláciu signálu $c_j(n)$ je rozdelenie na párne a nepárne vzorky veľmi výhodné. V bloku predikátora (P) sa používa postupnosť párnych vzoriek (resp. koeficientov) na predikciu postupnosti nepárnych vzoriek. Cieľom je získanie čo nejmenších hodnôt $d_{j-1}(n)$ v hornopriepustnej (HP) časti priamej DWT. Vo všeobecnosti operátor P závisí od špecifikácie konkrétnej banky filtrov.dvojkanálového systému analýza/syntéza [12]. Výstupný signál bloku korektora (K) mení hodnoty párnych vzoriek na základe nepárnych tak, aby čo najvernejšie

odzrkadľovali vlastnosti vstupného signálu $c_j(n)$, t.j. aby predikcia bola účinná aj pri ďalších krokoch.



Obr. 1 Lifting metóda imlementácie priamej DWT pre jeden stupeň rozkladu signálu $c_j(n)$ na signál aproximácie $c_{j-1}(n)$ a signál detailu $d_{j-1}(n)$.

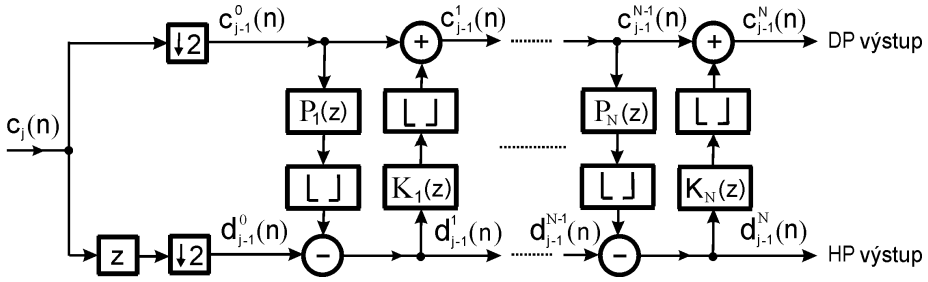
Predpokladajme, že vstupný diskrétny signál $x(n) = c_j(n)$, potom úplnú LI implementáciu priamej DWT tohto signálu dostaneme aplikovaním jedného stupňa LI rekurzívnym spôsobom ako to vidno na Obr.2.



Obr. 2 Úplna implementácia priamej DWT : rekurzívne rozdelenie na aproximácie a detaily.

Po tejto priamej DWT dostaneme j detailových postupností, t.j. $d_{j-1}(n)$ až d_0 , každá postupne s $n=2^{j-1}$ až 2^0 koeficientami a jednu aproximačnú postupnosť $c_0(n)$, $n= 2^0=1$, t.j. iba s jedným koeficientom.

Každá DWT založená na lifting implementácii môže byť ľahko modifikovaná na celočíselnú DWT (CČDWT), pridaním operácií zaokrúhľovania na úkor zavedenia nelinearity do tejto transformácie. Takéto celočíselné transformácie sú reveribilné. Implemetácia transformovanej celočíselnej DWT v zovšeobecnenom tvare, ktorá sa získa Z-transformáciou jej časovej lifting implemetácie (LI) je na Obr. 3. V hornej vetve za decimátorom s faktorom decimácie 2 dostávame párne vzorky a v dolnej vetve za kaskádnym zapojením bloku z a dacimátora dostávame nepárne vzorky vstupného diskrétneho signálu $c_j(n)$.



Obr. 3 Zovšeobecnená bloková transformovaná LI j-tého stupňa priamej CČDWT.

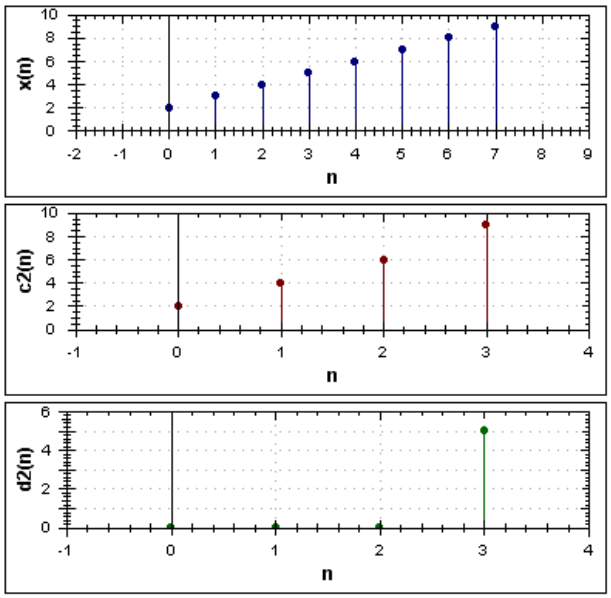
V každom kroku lifting *implementácie DWT sú zaokrúhlené signály z predikčného $P_i(z)$ a korekčného filtra $K_i(z)$, i=1,…,N* na celé čísla. Potom môže byť použitá celočíselná aritmetika, keďže aj vstupné údaje aj koeficienty týchto filtrov sú vyjadrené pomocou celočíselných hodnôt. Potom vzťahy pre j-tý stupeň priamej CČDWT v časovej oblasti budú [12]

$$d_{j-1}^{i}(n) = d_{j-1}^{i-1}(n) - \left\lfloor \sum_{r=0}^{s-1} p_i(r) c_{j-1}^{i-1}(n+r) + 1/2 \right\rfloor$$
$$c_{j-1}^{i}(n) = c_{j-1}^{i-1}(n) + \left\lfloor \sum_{r=0}^{s-1} k_i(r) d_{j-1}^{i-1}(n-r) + 1/2 \right\rfloor \quad (1)$$

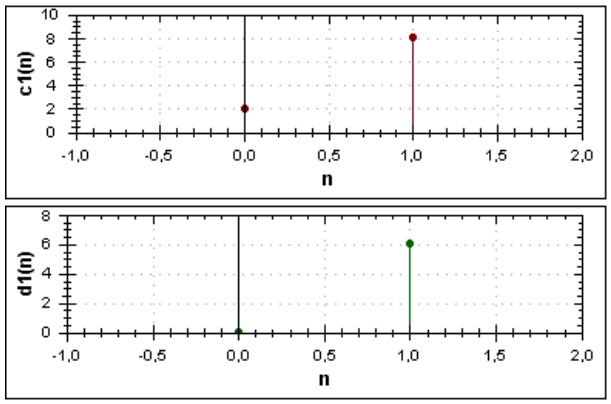
kde s je rád prediktora, resp. korektora a operácia zaokrúhlenia je označená Gaussovými zátvorkami $\lfloor\ \rfloor$, pričom výraz $\lfloor a \rfloor$ predstavuje najväčšie celé číslo menšie alebo rovné a. Vychádzajúc z rov.(1) možno ukázať pre biortogonálnu (5,3) CČDWT, ktorej rád s = 2 a i =1, že platí $p_1(1) = p_1(0) = p_1 = 1/2$ *a* $k_1(0) = k_1(1) = k_1 = 1/4$.

III. EXPERIMENTÁLNE VÝSLEDKY

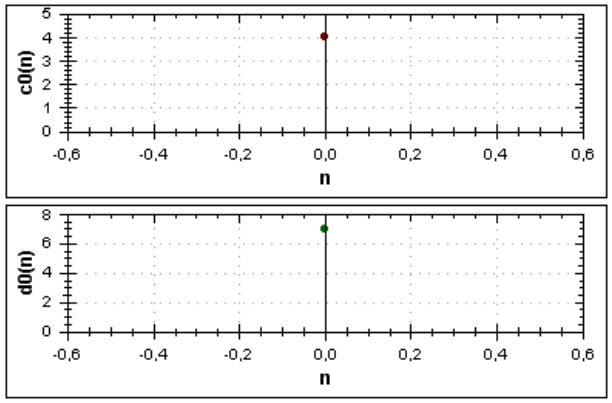
V nasledujúcich experimentoch vykonáme priamu biortogonálnu (5,3) CČDWT a výsledky zobrazíme dvomi spôsobmi. Ako vstupný signál sme zvolili v prvom prípade lineárnu postupnosť $x(n) = c_3 = \{2,3,4,5,6,7,8,9\}$ s počtom vzoriek 8. Na Obr. 4a je zobrazený vstupný signál $x(n)$ spolu so štyrmi aproximačnými $c_2(n)$ (červená) a štyrmi detailovými $d_2(n)$ (zelená) koeficientami po prvom stupni transformácie, pričom prvé tri detailové koeficienty sú nulové.



(a)



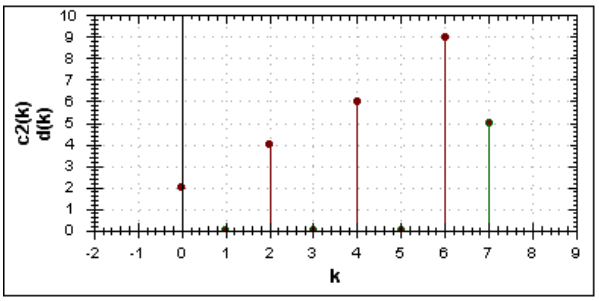
(b)



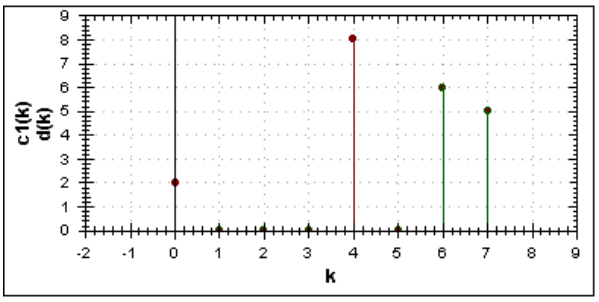
(c)

Obr. 4 Priama (5,3) CČDWT (a) prvý stupeň, 4 aproximačné $c_2(n)$ a detailové koeficienty $d_2(n)$, (b) druhý stupeň, 2 aproximačné $c_1(n)$ a detailové koeficienty $d_1(n)$, (c) tretí stupeň, 1 aproximačný $c_0(n)$ a detailový koeficient $d_0(n)$.

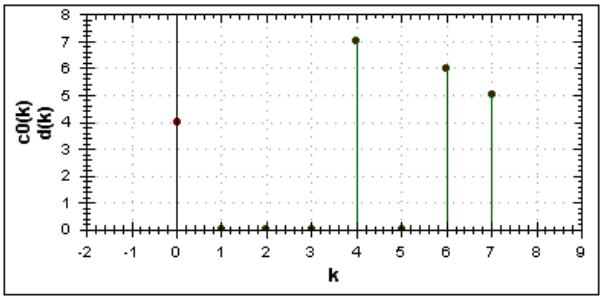
Na Obr. 4b a 4c je výsledok po druhom a treťom stupni transformácie. Na Obr. 5 sú zobrazené aproximačné a detailové koeficienty pre (5,3) CČDWT druhým spôsobom (nahradzovanie aproximačných koeficientov z predchádzajúceho stupňa rozkladu koeficientami detailov v nasledujúcom stupni) pre ten istý vstupný signál. Porovnaním grafu zobrazujúceho vstupný signal *x(n)* na Obr. 4a a grafu na Obr. 5c, na ktorom je zobrazený výsledok po úplnej transformácii vidíme, že hodnoty vstupného signálu sú nahradené v tej istej pamäti jedným aproximačným koeficientom a siedmymi detailovými koeficientami.



(a)



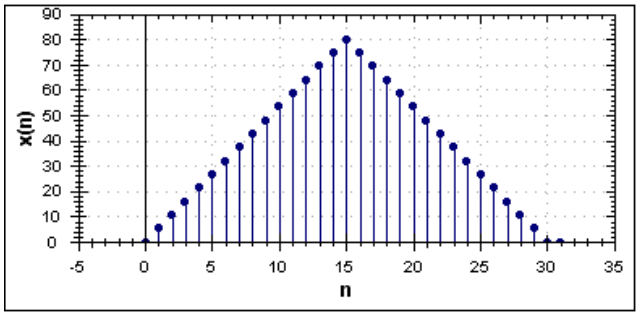
(b)



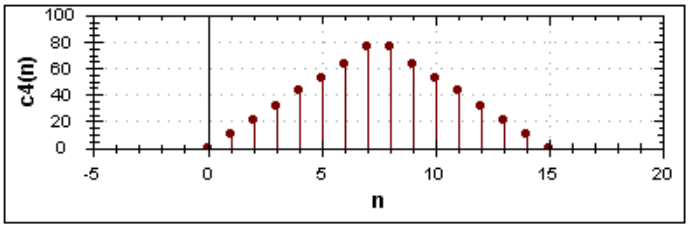
(c)

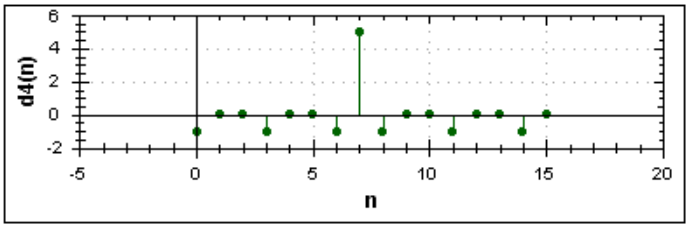
Obr. 5 Priama (5,3) CČDWT (a) prvý stupeň, (b) druhý stupeň, (c) tretí stupeň.

Toto je dôležitá vlastnosť lifting metódy DWT pri šetrení pamäťou. Inými slovami povedané, nie je potrebná žiadna dodatočná pamäť a pôvodný signál môže byť nahradený jeho waveletovou reprezentáciou. Táto vlastnosť je označovaná v anglickej literatúre ako in-place. Na Obr. 6a je zobrazený vstupný signál *x(n)* trojuholníkového tvaru s počtom 32 vzoriek.

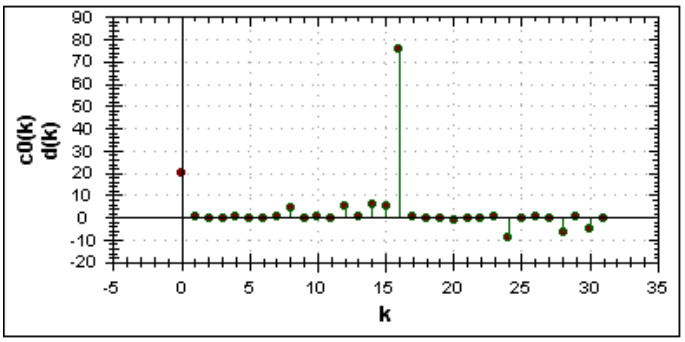


(a)





(b)



(c)

Obr. 6 Priama (5,3) CČDWT (a) vstupný signál (b) prvý stupeň, 16 aproximačných $c_4(n)$ a detailových $d_4(n)$koeficientov (c) úplná - druhý spôsobom zobrazenia.

Na Obr. 6b sú zobrazené aproximačné $c_4(n)$ a detailové $d_4(n)$ koeficienty po prvom stupni transformácie a na Obr. 6c je výsledok po úplnej transformácii zobrazený druhým spôsobom (in-place).

IV. ZÁVER

V článku je prezentovaná lifting implementácia diskrétnej waveletovej transformácie a prechod k zovšeobecnenej implementácii celočíselnej diskrétnej waveletovej transformácie zavedením operácie zaokrúhlenia, na úkor zavedenia nelinearity do tejto transformácie. Aj napriek tomu sú celočíselné transformácie reverzibilné. Výsledky experimentov poukazujú na dobrú dekorelačnú schopnosť biortogonálnej (5,3) CČDWT. Celočíselné údaje sú veľmi dôležité a užitočné napr. pre bezstratové kódovanie signálov.

LITERATÚRA

[1] I. Daubechies, Ten Lectures on Wavelets. Society for Industrial and Applied Mathematics (SIAM), CBMS Series, April 1992.
[2] M. Holschneidr, Wavelets. An Analysis Tool. Express, Oxford, New York, 1975.
[3] N. L. Fliege, Multirate Digital Signal Processing: Multirate Systems, Filter Banks, Wavelets. West Sussex England John Willey & Sons, 1999.
[4] G. Strang -T. Nguyen, Wavelets and Filter Banks. Wellesley-Cambridge Press, 1996.
[5] P. P. Vadyanathan, Multirate Systems and Filter Banks, Pretice Hall, 1993
[6] S. Malat, A Wavelet Tour of Signal Processing (2nd Ed.). Acadenic Press, UK, 2001.
[7] R. Vích - Z. Smékal, Číslicové filtry. Academia, nakl. AV ČR, Praha 2000.
[8] I. Daubechies - W. Sweldens, Factoring Wavelet Transforms into Lifting Steps. J. Fourier Anal. Appl., Vol. 4, No. 3, 1998, pp. 247-269.
[9] W. Sweldens, The Lifting Schneme: A New Philosophy in Biorthogonal Wavelet Constructions. Proc.SPIE 2569, 1995, pp. 68-79.
[10] J. Mihalík - J. Zavacký - J. Dzivý, Perfect Reconstruction 2DQMF Bank for Subband Coding. Journal Electrical Engineering, Vol. 47, No. 7-8, 1996, pp. 195-201.
[11] A. R. Calderbank - I. Daubechies - W. Sweldens - B. L. Yeo, Wavelet transforms that Map Integers to Integers. Applied Computational Harmonic Analysis, Vol. 5, No. 3, 1998, pp.332-369.
[12] J. Mihalík - J. Zavacký, Diskrétne spracovanie signálov, LČSOV FEI TU Košice, 2011, ISBN 978-80-553-0730-5, (319 strán).

Zmena vzorkovacej frekvencie s neceločíselným faktorom

Jozef ZAVACKÝ

Katedra elektroniky a multimediálnych telekomunikácií, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

jozef.zavacky@tuke.sk

***Abstrakt* – Tento článok sa venuje zmene vzorkovacej frekvencie s neceločíselným faktorom L/D. Zmena vzorkovacej frekvencie sa dosiahneme kaskádnym zapojením interpolátora s celočíselným faktorom L, dolnopriepustného číslicového filtra s konečnou impulznou odozvou a decimátora s celočíselným faktorom D. Faktory L a D musia byť nesúdeliteľné celé čísla. Úloha dolnopriepustného číslicového filtra s prenosovou funkcciou G(z) pri efektívnej neceločíselnej konverzii vzorkovacej frekvencie je dvojaká, pôsobí ako interpolačný filter, ktorý potláča efekt vytvárania kópii a súčasne ako antialiasingový filter. Pre dobré odstránenie kópii, musí byť medzná frekvencia nepriepustného pásma dolnopriepustného filtra G(z) nižšia ako π/L a zabránenie aliasingu vyžaduje, aby medzná frekvencia nepriepustného pásma bola nižšia ako π/D. Vykonali sme experimenty pre prevzorkovanie so zvýšením vzorkovacej frekvencie pre dva signály s ohraničeným frekvenčnými spektrom s použitím funkcií v prostredí Matlab.**

***Kľúčové slová* — decimátor, interpolátor, prevzorkovanie, neceločíselný faktor, dolnopriepustný číslicový filter, mnohorýchlostné systémy**

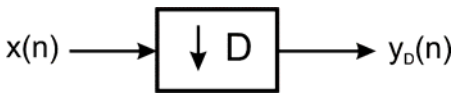
I. Úvod

Mnoho diskrétnych systémov so zmenou vzorkovacej frekvencie našlo svoje uplatnenie medzi modernými algoritmami pre číslicové spracovanie signálov vzhľadom k svojej vysokej výpočtovej efektívnosti zvlášť v spojení s implementáciou na signálovom procesore [1-8]. Takéto systémy pracujúce s viacerými vzorkovacími frekvenciami sú známe ako mnohorýchlostné systémy (Multirate Systems). Svoje uplatnenie nachádzajú v mnohých oblastiach ako sú komunikačné systémy, spracovanie reči, simulácia akustiky miestnosti využívajúca wavelety, kompresia obrazu, anténne systémy, radarové systémy, systémy utajovania hovorov, transmultiplexory, senzorové systémy, modely pre návrh architektúry mikroelektronických obvodov atď. Typickým príkladom využitia zmeny vzorkovacej frekvencie je konverzia hudobných záznamov, spracovaných v hudobných štúdiách (vzorkovacia frekvencia býva f_v=48 kHz alebo f_v=96 kHz) a ich ukladanie na kompaktný disk (CD), kde sa používa vzorkovacia frekvencia f_v=44,1 kHz. Riešenie by mohlo byť zdánlivo jednoduché. Digitálny záznam by sme previedli spätne na analógový signál a ten by sme potom znovu vzorkovali s novou vzokovacou frekvenciou. Pri analógovo číslicovom (A/Č) a číslicovo analógovom (Č/A) prevode môžu byť zavedené také závažné chyby a rušenia, že po vykonaní tohoto postupu je výsledný signál takmer nepoužiteľný. Ďaleko výhodnejšie je zmeniť vzorkovaciu frekvenciu priamo u diskrétneho alebo číslicového signálu. Dve základné operácie zmeny vzorkovacej frekvencie sú decimácia a interpolácia.

II. Decimátor a Interpolátor

Decimátor s celočíselným faktorom D, ktorého bloková schéma je na Obr.1 je charakterizovaný vstupno-výstupným vzťahom

$$y_D(n) = x(nD) \tag{1}$$



Obr. 1 Decimátor s faktorom *D*.

pričom pre jednoduchosť uvažujeme jednotkovú diskretizačnú periódu T=1. To znamená, že z pôvodnej postupnosti vzoriek $x(n)$ sa vyberie každá D-tá vzorka do postupnosti $y_D(n)$. Alebo ináč povedané výstupný signál v čase n je rovný vstupnému signálu v čase nD. Z-transformácia signálu $y_D(n)$ na výstupe decimátora je daná výrazom [9-10]

$$Y_D(z) = \frac{1}{D}\sum_{m=0}^{D-1} X(z^{1/D} W^m) \tag{2}$$

kde $W = e^{-j2\pi/D}$. Potom spektrum decimovaného signálu $y_D(n)$ dostaneme po substitúcií $z = e^{j\omega}$ v rov. (2) výsledkom čoho je

$$Y_D(e^{j\omega}) = \frac{1}{D}\sum_{m=0}^{D-1} X(e^{j(\omega-2\pi n)/D}) \tag{3}$$

Pretože decimácii v časovej oblasti zodpovedá kompresia, to má za následok expanziu vo frekvenčnej oblasti. Člen s $m = 0$ je D-násobne natiahnutou verziou spektra $X(e^{j\omega})$, pričom D-1 členov radu (3) pre m > 0 sú posunuté verzie nultého člena. Ak vstupný diskrétny signál $x(n)$ nemá ohraničené spektrum na oblasti $|\omega| < \pi/D$, potom decimácia zapríčiňuje prekrývanie spektrálnych členov (aliasing). Vtedy spektrálne členy s $m > 0$ sa budú prekrývať so základným členom spektra pre $m = 0$ v oblasti $0 \le \omega \le \pi$.

Bloková schéma interpolátora s faktorom L je na Obr.2 a je charakterizovaný vstupno-výstupným vzťahom

$$y_I(n) = \begin{cases} x(n/L) & ak\ n\ je\ celistvým\ násobkom\ L \\ 0 & ináč \end{cases} \tag{4}$$

čo znamená, že interpolátor vloží (L-1) nulových vzoriek medzi susedné vzorky vstupného signálu $x(n)$.



Obr. 2 Interpolátor s faktorom L.

Potom Z-transformácia signálu na výstupe interpolátora $y_I(n)$ je daná výrazom

$$Y_I(z) = \sum_{n=0}^{\infty} x(n/L) z^{-n} = \sum_{n=0}^{\infty} x(n) z^{-Ln} = X(z^L) \tag{5}$$

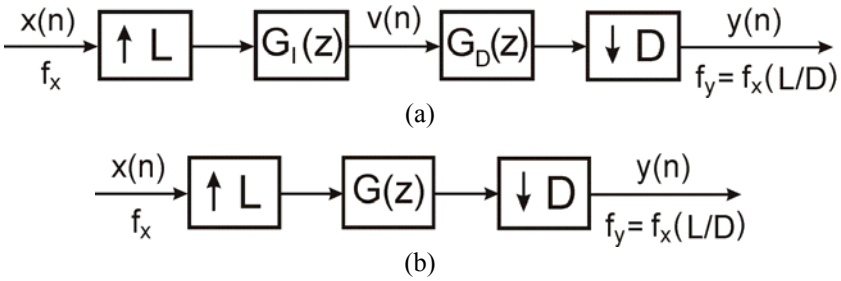
a jeh komplexné frekvenčné spektrum

$$Y_I(e^{j\omega}) = X(e^{j\omega L}) \tag{6}$$

t.j. $Y_I(e^{j\omega})$ je L-násobne stlačenou verziou $X(e^{j\omega})$. Mnohonásobne objavujúce sa kópie v spektre $Y_I(e^{j\omega})$ sú vytvárané interpolátorom a tento jav je známy ako efekt vytvárania kópií (imaging effect). Tieto kópie , ak je to potrebné, môžu byť potlačené DP filtrom.

III. Zmena Vzorkovacej Frekvencie Diskrétneho Signálu s Neceločíselným Faktorom

Zmena vzorkovacej frekvencie s racionálnym faktorom L/D nazývaná aj prevzorkovanie sa dosiahne najprv zvýšením vzorkovacej frekvencie pomocou interpolátora s faktorom L a potom znížením pomocou decimátora s faktorom D. Teda zmenu vzorkovacej frekvencie s racionálnym faktorom L/D dosiahneme kaskádnym zapojením interpolátora a decimátora ako to vidno na Obr. 3a. Faktory L a D musia byť nesúdeliteľné celé čísla, to znamená že neexistuje žiady celočíselný spoločný násobok medzi číslami L a D.



Obr. 3 Zmena vzorkovacej frekvencie s nesceločíselným faktorom *L/D* (a) všeobecná bloková schema, (b) výsledná bloková schema.

Pri implementácii podľa blokovej schémy na Obr. 3a sa najprv interpoluje pôvodný signál *x(n)* a potom sa filtruje dolnopriepustným interpolačným filtrom s prenosovou funkciou $G_I(z)$. Interpolovaný signál *v(n)* je ďalej filtrovaný pomocou dolnopriepustného antialiasingového filtra s prenosovou funkciou $G_D(z)$ a nakoniec je decimovaný s faktorom *D*. Vzorkovacia frekvencia výstupného signálu *y(n)* je *L/D* násobok vzorkovacej frekvencie pôvodného signálu *x(n)*. Pretože interpolačný filter $G_I(z)$ a decimačný filter $G_D(z)$ pracujú s rovnakou vzorkovacou frekvenciou, môžu byť nahradené jediným dolnopriepustným filtrom s prenosovou funkciou *G(z)* ako to znazorňuje obr.3b. Dolnopriepustný (DP) filter *G(z)* musí byť navrhnutý tak, aby odstránil efekt vytvárania kópii, ktoré sa vytvárajú pri interpolácii a odstránil aliasing vznikajúci pri decimácii. Vďaka vhodne navrhnutému filtru *G(z)* s končnou impulznou odozvou (KIO) môže byť neceločíselná zmena vzorkovacej frekvencie realizovaná pomocou výpočtovo efektívnej štruktúry tak ako je to vidno na Obr. 3b. Úloha filtra *G(z)* pri efektívnej neceločíselnej konverzii vzorkovacej frekvencie podľa Obr. 3b je dvojaká: pôsobí ako interpolačný filter $G_I(z)$, ktorý potláča efekt vytvárania kópii a súčasne ako antialiasingový filter $G_D(z)$. Pre dobré odstránenie kópii, musí byť medzná frekvencia nepriepustného pásma dolnopriepustného filtra *G(z)* nižšia ako π/L a zabránenie aliasingu vyžaduje, aby medzná frekvenciou nepriepustného pásma bola nižšia ako π/D. Preto medzná frekvencia nepriepustného pásma pre DP filter *G(z)* je určená takto

$$\omega_s = \min\left(\frac{\pi}{L}, \frac{\pi}{D}\right) \tag{7}$$

Ideálna amplitúdovou frekvenčnou charakteristikou filtra *G(z)* bude

$$G(\omega) = \begin{cases} L, & |\omega| \le \min\left(\frac{\pi}{L}, \frac{\pi}{D}\right) \\ 0, & inde \end{cases} \tag{8}$$

Pre veľké hodnoty faktorov *L* a *D* sa požadujú filtre s veľmi úzkymi priepustnými pásmami. V Matlabe existuje niekoľko funkcií, ktoré vykonávajú neceločíselnú zmenu vzorkovacej frekvencie tak ako to je zobrazená na Obr. 3b.

Je zrejmé, že DP filter *G(z)* v blokovej schéme na Obr.3b aproximuje s určitou chybou ideálnu amplitúdovú frekvenčnú charakteristiku danú rov.(8) a v dôsledku toho proces prevzorkovania spôsobí skreslenie v spektre prevzorkovaného signálu. Odvodíme vzťah pre spektrum prevzorkovaného signálu $Y(e^{j\omega})$ konvertora na Obr. 3b, pomocou spektra $X(e^{j\omega})$ vstupného diskrétneho signálu, komplexnej frekvenčnej charakteristiky $G(e^{j\omega})$ DP filtra a faktorov *L* a *D*. Príslušný vzťah odvodíme v oblasti Z-transformácie a potom urobíme substitúciu $z = e^{j\omega}$.

Z-transformáciu prevzorkovaného signálu získame v troch krokoch

1. Vykonaním interpolácie s faktorom *L*, podľa rov.(6), Z- transformácia interpolovaného signálu je $X(z^L)$.
2. Filtráciou interpolovaného signálu DP filtrom s prenosovou funkciou G(z), výsledkom čoho je súčin $X(z^L).G(z)$.
3. Posledný krok je decimácia s faktorom *D*. Použitím rov.(2) pre Z-transformáciu interpolovaného a filtrovaného signálu dostávame

$$Y(z) = \frac{1}{D}\sum_{k=0}^{D-1} X(z^{L/D}W_D^{-k})G(z^{1/D}W_D^{-k}) \tag{9}$$

Po substitúcii $z = e^{j\omega}$ v rov.(9) dostaneme požadovaný vzťah pre komplexné frekvenčné spektrum prevzorkovaného signálu

$$Y(e^{j\omega}) = \frac{1}{D}\sum_{k=0}^{D-1} X(e^{j(L\omega-2\pi k)/D})G(e^{j(\omega-2\pi k)/D}) \tag{10}$$

Ak použijeme ideálnu komplexnú frekvenčnú charakteristiku $G(e^{j\omega})$ danú rov.(8), potom komplexné frekvenčné spektrum

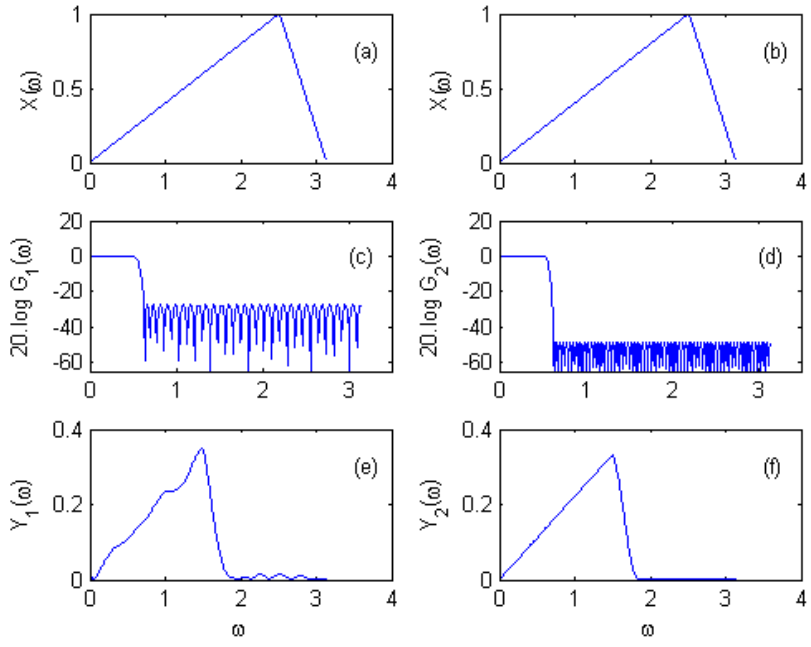
$$Y(e^{j\omega}) = \frac{1}{D}X(e^{jL\omega/D}) \tag{11}$$

Spektrum $Y(e^{j\omega})$ je rovné spektru vstupného signálu $X(e^{j\omega})$, ktoré je stlačené s faktorom L a roztiahnuté s faktorom D, navyše vynásobené s $1/D$. Je zrejmé, že celkové vlastnosti konvertora závisia od vlastností DP filtra.

IV. Experimentálne Výsledky

V nasledujúcom experimente budeme demonštrovať vplyv vlastností pásma tlmenia na spektrum prevzorkovaného signálu. Uvažujeme konvertor s faktormi L=5 a D=3. Ako vstupný signál $x(n)$ si zvolíme signál s ohraničeným spektrom trojuholníkového tvaru. Takýto signál môžeme v Matlabe vygenerovať pomocou funkcie *fir2, x = fir2(256,[0,0.8,1], [0,1,0])*. Navrhneme dva rôzne filtre s konečnou impulznou odpoveďou (KIO), aby sme porovnali vplyv dvoch rôznych účinkov pásmiem tlmenia na spektrum prevzorkovaného signálu. Pomocou rov.(8) určíme medznú frekvenciu ω_s nepriepustného pásma pre obidva filtre $\omega_s = min(\pi/3, \pi/5)=\pi/5$. Pre medznú frekvenciu priepustného pásma zvolíme $\omega_p=\pi/6$. Matlabovská funkcia *firgr* vypočíta impulzné charakteristiky $g_1(n)$, $g_2(n)$ dvoch KIO filtrov s dĺžkami $N_1 = 71$ *a* $N_2 = 151$,*t.j. g1 = firgr(70,[0,1/6,1/5,1],[1,1,0,0])* *a g2 = firgr(150,[0,1/6,1/5,1],[1,1,0,0])*.Funkcia *upfirdn* vykoná konverziu vzorkovacej frekvencie vstupného signálu x pre L = 5 a D = 3, *y1 = upfirdn(x,g1,5,3)* *a y2 = upfirdn(x,g2,5,3)*.

Na Obr. 4 je zobrazené amplitúdové spektrum $X(\omega)$ vstupného diskrétneho signálu, amplitúdové frekvenčné charakteristiky $G_1(\omega)$,$G_2(\omega)$ v dB obidvoch filtrov, a odpovedajúce modulové spektrá $Y_1(\omega)$ a $Y_2(\omega)$ prevzorkovaných signálov. Ako vidno na Obr.4(c) a (e) nepriepustne pásmo filtra $G_1(z)$ s tlmením 30 dB nedokáže potlačiť aliasing a odstrániť kópie zo spektra prevzorkovaného signálu. Na Obr.4 (d) a (f) môžeme vidieť, že nepriepustné pásmo filtra $G_2(z)$ s tlmením 50 dB odstráni viditeľné účinky aliasingu a vytvárania kópií zo spektra signálu.

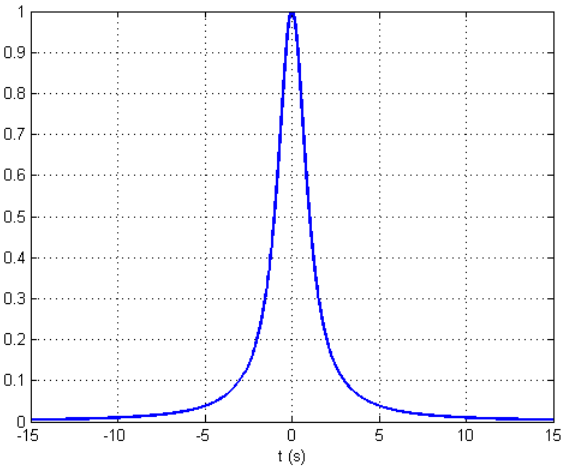


Obr. 4 Prevzorkovanie s L/D = 5/3: (a) a (b) Amplitúdové spektrum $X(\omega)$ vstupného signálu, (c) Amplitúdová frekvenčná charakteristika $20log\ G_1(\omega)$, (d) Amplitúdová frekvenčná charakteristika $20log\ G_2(\omega)$, (e) Amplitúdové spektrum prevzorkovaného signálu $Y_1(\omega)$, (f) Amplitúdové spektrum prevzorkovaného signálu $Y_2(\omega)$.

Druhý experiment sme vykonali so signálom (Obr. 5)

$$x(t) = \frac{1}{1+t^2}(1 - e^{-\omega_m}.\cos\,\omega_m t + t.e^{-\omega_m}.\sin\,\omega_m t)\,, \quad t \in \langle -\infty, \infty \rangle \tag{12}$$

ktorého spektrum nech je ohraničené maximálnou kruhovou frekvenciou ω_m=10 (rad.s^{-1}).



Obr. 5 Spojitý signál x(t).

Kruhovú vzorkovaciu frekvenciu pre tento signál volíme ω_x = 30 (rad.s^{-1}). Vstupný diskrény signál $x(nT_x)$ dostaneme vzorkovaním spojitého signálu $x(t)$ daného rov.(12) so vzorkovacou periódou $T_x = 2\pi/\omega_x$ = 0.2094395102 (s), resp. vzorkovacou frekvenciou f_x = 4.774648293(Hz). Vykonali sme prevzorkovanie signálu $x(nT_x)$ pre pomery L/D : 3/2, 5/3, 7/4, 9/5, 14/3, 19/3, 28/3, 37/2, 160/147 pomocou Matlabovskej funkcie y = *resample(x,L,D)*. Pre návrh DP filtra s KIO pri aplikácii funkcie *resample* je použitá funkcia *firls* s Kaiserovým oknom s parametrom *beta* 5 a dĺžkou filtra N = 11. Vypočítali sme strednú kvadratickú chybu

$$\varepsilon = \frac{1}{M}\sum_{n=0}^{M}\left(x(nT_y) - \hat{x}(nT_y)\right)^2 \tag{13}$$

kde $\hat{x}(nT_y)$ je prevzorkovaný signál, $x(nT_y)$ je diskrétny signál získaný vzorkovaním pôvodného spojitého signálu s novou vzorkovacou periódou T_y a M je počet vzoriek. Získané výsledky sú uvedené v Tabuľke 1.

Tabuľka 1.

L/D	f_y	T_y	***M***	***ε***
3/2	7.161972442	0.1396263401	215	2.1188e-8
5/3	7.957747157	0.1256637061	239	2.1159e-8
7/4	8.355634515	0.1196797201	251	2.1117e-8
9/5	8.594366932	0.1163552834	258	2.1339e-8
17/4	20.29225525	0.04927988475	609	2.1510e-8
19/3	40.31925224	0.03306939635	908	2.2138e-8
28/3	44.56338408	0.02243994752	1337	2.1646e-8
37/3	58.88732895	0.01698158191	1767	2.1796e-8
160/147	5.196896102	0.19242255	156	2.1554e-8

V. ZÁVER

V tomto článku sme navrhli spôsob zmeny vzorkovcej frekvencie diskrétneho signálu implementáciou štruktúry kaskádneho zapojením interpolátora s celočíselným faktorom L, dolnopriepustného číslicového filtra s konečnou impulznou odozvou a decimátora s celočíselným faktorom D. V prvom experimente sme ukázáli vo frekvenčnej oblasti ako vlastnosti konvertora závisia od vlastností DP filtra, čo sme demonštrovali pre zvolený vstupný signál trojuholníkového tvaru s ohraničeným frekvenčným spektrom. V druhom experimente sme vykonali zvýšenie vzorkovacej frekvencie pre iný typ vstupného signálu taktiež s ohraničeným frekvenčným spektrom s rôznymi s neceločíselnými faktoromi L/D a vypočítali sme strednú kvadratickú chybu ε. Dosiahnuté výsledky potvrdili vysokú presnosť vykonanej konverzie kedže stredná kvadratická chyba je veľmi blízka nule a to dokonca aj pre veľký pomer L/D = 160/147. Ďalší výskum v tejto oblasti je možné zamerať na polyfázovú implementáciu konvertorov s neceločíselnými faktormi L/D a na konverziu vzorkovacej frekvencie s ľubovoľným faktorom.

LITERATÚRA

[1] J. Zavacký - J. Mihalík - I. Gladišová, Implementácia diskrétnej waveletovej transformácie v štandarde JPEG-2000. Slaboproudý obzor,roč. 63, č. 3-4, 2007, s. 5-9, ISSN 0037-668X.
[2] N. J. Fliege, Multirate Digital Signal Processing: Multirate Systems, Filter Banks, Wavelets. John Willy & Sons, Chichester, 1994.
[3] R. E. Crochiere - L. R. Rabiner, Multirate Digital Signal Processing, Englewood Cliffs, NJ : Prentice- Hall, 1983.
[4] G. Jonvanović-Doleček, Multirate Systems: Design & Applications. London , U. K. :Idea Group Publshing, 2002.
[5] H. S. Malvar, A Link between Transform Coding and Multirate Filter Banks. IEEE Int. Symp. Circuit and Systems, 1988, pp. 835-838.
[6] W. M. Campbel - T. W. Parks, Design of a Class of Multirate Systems. In: ICASSP Proceedings, 1995, pp. 1308-1311.
[7] F. J. Harris, Multirate Signal Processing for Communications. Upper Saddle River, NJ: Prentice – Hall, 2004
[8] M. Laddomaada, Dessign of Multistage Decimation Filters Using Czclotonic Polznomials : Optimization and Design Issues, IEEE Trans. Circuit Syst. I, Vol. 55, No. 7, 2008, pp. 1977-1987.
[9] J. Mihalík - J. Zavacký, Diskrétne spracovanie signálov, LČSOV FEI TU Košice, 2011, ISBN 978-80-553-0730-5, (319 strán).
[10] J. Mihalík - J. Zavacký - I. Gladišová, Mnohokanálové diskrétne sústavy. Slaboproudý obzor,roč. 65, č.4, 2009, str. P1-P8, ISSN 0037-668X.

Design of Solution for Video Streaming of Interactive Whiteboard Surface

Dávid CYMBALÁK, Miroslav MICHALKO

Department of Computers and Informatics, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Slovak Republic

david.cymbalak@tuke.sk, miroslav.michalko@cnl.sk

***Abstract* — This paper deals with a design of multimedia solution, which is usable for streaming video content from interactive whiteboard surface. Based on the analysis of existing streaming technologies, chosen streaming components in cooperation with interactive whiteboard are interconnected into a interactive solution with educational abilities. The overall concept allows capturing projected image on interactive whiteboard with layer of drown notes on whiteboard surface. This video input is combined with audio from microphone on lecturer body. Captured content is encoded in proper bit-rate depending on the resolution of projected image on whiteboard. Encoded video is streamed via media server to flash player located in web interface, where the lecture could be watched by students. It is also possible to archive streamed lecture in form of video file. Experimental realization of proposed design was made by creation of testing web page with ability to watch streamed broadcast of experimental testing lecture.**

***Keywords* — Interactive whiteboard, Video streaming, E-learning**

I. INTRODUCTION

In conjunction with the concept of modern technologies in education it is increasingly talking about the using of interactive whiteboard. Imaging technique facilitates the teacher's work in form of expanding the possibilities for displaying visual materials, which enhance the diversity of learning. Technology of interactive whiteboard allows dynamically controlling of presented text or graphical information and it brings the interactive elements through the original materials. Usability of whiteboards is currently limited to teach students located in classroom, but the image from the whiteboard's surface could become perfect source for video streaming via network to remote students.

The Technical University of Kosice has an adequate amount of interactive whiteboards, so there is place to integrate this element into learning process and to provide appropriate trainings to teachers. In present these interactive whiteboards are used by teachers to present materials for lectures, but whiteboard surface screens are not streamed over the network, so the interaction in lecture is limited for students sitting directly in the class. Some lectures are captured by camera and streamed live or recorded to archive, but there is a lack of picture quality of presented materials which leads to decrease of readability. For those reasons it is necessary to interconnect mentioned learning elements and develop multimedia solution which will satisfy most of requirements in modern educational institution.

II. ANALYSES OF EXISTING TECHNOLOGIES

A. Interactive Whiteboards

Interactive whiteboard is a large electronic projection area, which is connected to computer. Projector is generally mounted with whiteboard on one stand. Surface of whiteboard reflects the images from computer. Using electronic pen, special stylus or only finger it is possible to easily control computer, or work directly with the board – draw lines, write notes. All movements realized by touching of surface while drawing are captured from whiteboards by sensors and are shown on surface in real time [1,2]. The advantage is the possibility to save notes, texts and images drawn on the surface. Most of the software cooperated with interactive whiteboards includes additional features such as handwriting recognition, rulers or protractors.

Connection of interactive whiteboard with computer is commonly provided by USB,

FireWire interface or WIFI. Transforming position on board's surface to position of cursor in computer's operating systems is ensured by system driver. Capturing the pen's position on surface is usually based on electromagnetic or optic principles. In this solution will be used interactive infrared whiteboard QOMO QWB200-PS with USB connection and infra red sensors located in the corners of the board, which allows controlling by pen and also by finger.

Unlike conventional projector, in this case are given higher requirements on the projector, which is mounted on whiteboard. The basic difference is minimized projection distance and maintaining a rectangular output image. Short-throw distance is needed to prevent the unwanted shadow from the person controlling the board. The current projector lens shines from top in acute angle, what would be causing a trapezoid shape of output image in conventional projectors. In this solution will be used projector NEC NP610SG.

B. Streaming Technologies

Streaming is a technology, which ensures the transfer of multimedia content over a network. It allows transmission of audio, video or even other multimedia content from sender to recipients. This technology can be used to real-time broadcasting, but also for receiving video on demand (VOD). Streaming media formats compared with conventional media formats brings many advantages. These benefits include: media content protection against piracy, controlling the flow of content, selection of allowed recipients or easy content management [3]. For this reason, especially due to affordability, these technologies are largely appearing as an appropriate means for providing educational services through the Internet. This all brings a new dimensions and more effective teaching methods then static text and images. Most common components of streaming technologies, which are usable for web based streaming solutions, are Wowza Media Server, Flash Media Live Encoder and JW player.

Wowza Media Server is fully interactive server for streaming multimedia content with full support for H.264. It is Java server platform, which is also suitable for high deployment with a large number of clients in the network and with continuous operation. There is also the possibility of using modules that provide control of dynamic content such as advertising or playlist. Streaming features are provided through an encrypted transfer with RTMP / RTMPS protocol and special security token, which generates a strong protection against applications for recording broadcast content. Broadcast reception is possible via the web interface, custom software applications, mobile devices or set-top box. Wowza is able to stream to all of these devices in parallel and selects the correct protocol and data flow depending on the client [4].

Flash Media Live Encoder is software, which is built to stream live video in real time and simultaneously convert to Flash. It supports codec H.264 and On2 VP6. The program can simultaneously live broadcast and save the broadcast to file in FLV format [4]. This feature is useful for creating video archives in IPTV solution. Encoder has various features influencing quality of streaming beginning from selecting the encoding resolution, ending with the choice of filters to determine the bit rate.

JWplayer is web based flash player with support of RTMP protocol and HTML5. It brings feature called "dynamic bit rate switching", which set the bit rate of video dynamically according the information from user about bandwidth. This feature decides what bit rate has to be provided to best ensure the continuity and quality of play. Playback can also stop and delay live streaming, then store streaming in temporary memory and then replay from memory for next watching. Player supports various extensions such as publish videos directly on social networks with one control or adding information directly to the video.

III. DESIGN OF SOLUTION

A. Expected Functionality

Expected functionality of proposed solution is delivering of live stream of presented materials with drawn notes on interactive whiteboard to students via network. Picture from whiteboard could by combined with audio from microphone capturing lecturer speech. Another expected functionality is providing of archive of historical lectures. For readability of presented materials it is important to provide the high definition resolution for all streaming services. Solution should provide possibility to interact with live streamed lecture in form of feedback messages to lecturer via web interface.

B. Interconnection between Chosen Components

Proposed interconnection between interactive whiteboard and mentioned chosen component of streaming technologies could be seen in Fig. 1

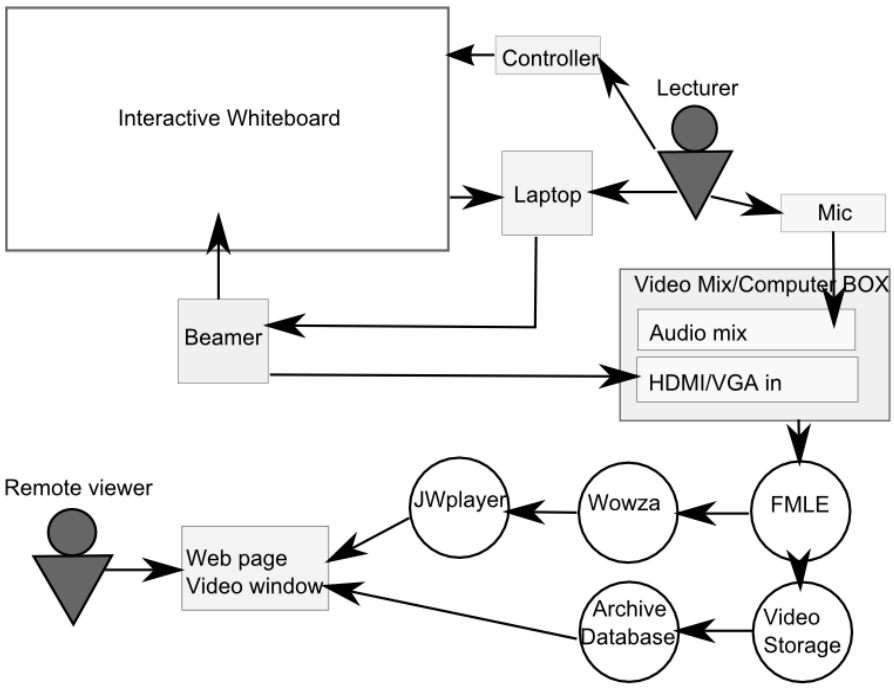


Fig. 1 Design of interconnection between components of proposed solution.

Lecturer is standing between interactive whiteboard and laptop. He is using controller or finger to drawing notes on whiteboard surface or controlling presentation. Coordinates of points of drown notes from surface and instruction for change position of visualized objects is send to laptop in cooperation with interactive whiteboard driver. Overlaid image consisting of presentations from laptop and drawings from whiteboard is distributed to beamer. The output is displayed on the interactive whiteboard on one hand; on second hand output is connected to HDMI input/ VGA input located in video streaming computer box or video mix device. Lector speech is captured by microphone, which is connected to audio mix together with audio output from lector's laptop. HDMI input devices containing raw resulting image from whiteboard surface is encoding via instance of Flash Media Live Encoder (FMLE). Encoded live stream is delivered as unicast to application of Wowza media server, which ensure optimized multicast streaming reachable from web based flash player – JW player. This window of JWplayer is located into code of experimental web page. Students are using personal computers with connection to testing web page and they have access to video player window, where it can be received video streams directly from interactive whiteboard with lector speech. Archiving of actual lecture is ensured by Flash Media Live Encoder in form streaming to file during live encoding. After ending of lecture, information about newly created file are stored into database.

C. Experimental Realization

Solution design was transformed to experimental realization with web interface. Testing Web page with JWplayer was created using the PHP, HTML, JavaScript language in cooperation with jQuery UI library and jQuery UI theme in CSS. Flash video player is resizable and address of streaming server can be changed dynamically. In this environment it is possible to integrate new component for extending the solution. For example there is possibility to add chat window in jQuery UI or another useful feature to improve interactive ability of solution.

In this experimental environment a couple of live broadcast from interactive whiteboard were realized. Couple of members and student of laboratory had rated the quality of streamed video at end device. It was streamed in various bitrate beginning from 500 kb/s to 5000 kb/s (Fig. 2). It was noticed, that streaming of presentation with drawn notes on interactive whiteboard had better results in bandwidth consumption than common video stream. In this case the average count of pixel changes is very small in each frame. The acceptable quality of video starts with bit-rate values from 500kb/s or 1000kb/s in case of situation, when presented slides are

changing very quickly or drawing on whiteboard is frequently used.

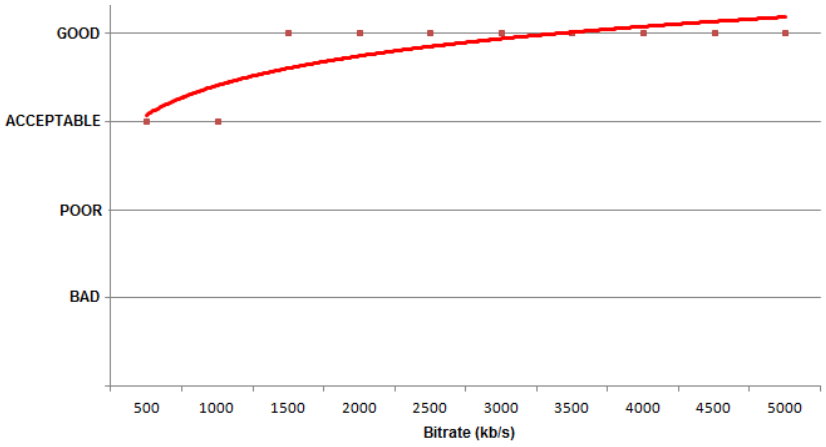


Fig. 2 Subjective quality of live streamed video containing PowerPoint presentation with layer of drawing notes on interactive whiteboard

IV. Conclusion

The idea of interconnection of interactive whiteboard with streaming technologies brings new perspective of modern education. Solution offers more interactivity between students and teacher opposed to standard streaming of lecture from camera source. Lecturer can use interactive whiteboard, where he can control presentation, drawing notes or using other integrated tools. Lecturer speech and also multilayered image from whiteboard is streaming via network to students via web interface. This functionality is very helpful for learning process not only for distance learning but also for organizing science conferences or corporate trainings. Future work on this solution will be concentrated on implementing comprehensive web portal, which will have not only described functionality, but also functionality such as voting systems, questions board, and ability to dynamically upload educational materials during live streaming.

Acknowledgment

This work is also the result of the project implementation Development of the Center of Information and Communication Technologies for Knowledge Systems (project number: 26220120030) supported by the Research & Development Operational Program funded by the ERDF.

References

[1] I. Baran, Interactive whiteboards in learning process, Thesis, Kosice : Technical University of Kosice, 2011.
[2] J. Dostal, Interactive whiteboards – significant contribution to education, 2009 Availible: http://www.ceskaskola.cz/2009/04/jiri-dostal-interaktivni-tabule.html
[3] O'Driscoll G., Next generation IPTV services and Technologies,Hoboken: Wiley, 2008.
[4] D. Cymbalak, Next generation IPTV, Diploma thesis, Kosice: Technical University of Kosice, 2011.

Category for Component-based Program System

Marián JENČIK, Daniel MIHÁLYI

Department of Computers and Informatics, Faculty of Electrical Engineering and Informatics Technical University of Košice, Slovak Republic

marian.jencik@student.tuke.sk, daniel.mihalyi@tuke.sk

Abstract **– In this contribution we describe basic terms of the component-based approach for program system construction by abstract behavioral type (ABT) view. ABT view represents new perspective for component-based program systems behavior because it allows system distribution on atomic software components and connectors, both modelled as abstract behavioral types, while each connector includes a well-defined inter-component communication protocol in the form of a relation.**

Keywords **– program component, connector, abstract behavioral type, component-based program system, inter-component communication protocol.**

I. Introduction

Component-based software engineering deals with construction of component-based program systems from existing program components, which connects together. [1].Program components are individually entities, can be used in different program systems and newly system which is composed from program components provides new functionality. We can say 'that the main idea of introducing a program component into software engineering is *reusability*.

Program component is defined as a non-trivial, independent and replaceable part of a system with contractually specified and explicitly described interfaces that satisfies clear function in the context of a well-defined program system [2].

Characteristics of a program component consists of the following:

- program component is a modular part of the system, that encapsulates its content;
- component operations are available only through consistent, well-published interfaces that satisfy communication standards;
- component interfaces include set of communication contact points through which a component have to be able to connect to other components;
- component can be implemented into program system independently of other components and implementation of the components can be realized by using different programming languages and paradigms provided that their composition and compatibility will be preserved;
- internal structure of a component may contain several other components;

In a programming paradigm where instances of components are their primitive buildings blocks, all decisions about which components will be used and together composed to build new program application must be made from outside what in generally means that a component cannot internally decide about the communication with another components.

II. Program System

Program system P is from the ABT view defined as component-based system, which consists of program component instances pk_i and connectors k_i (*glue code*) between them, which are together modelled (encapsulated) as *abstract behavior types (ABTs)*.

Program system P can be denoted as :

$$P = \left\{ pk_1 \overset{k_1}{\circ} pk_2 \overset{k_2}{\circ} \dots \overset{k_{n-1}}{\circ} pk_n \right\}. \tag{1}$$

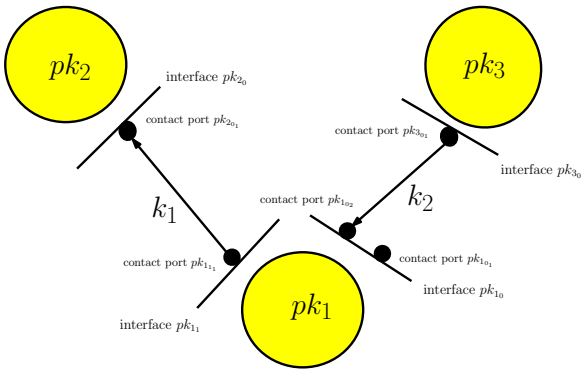


Fig. 1 Program system consists of three program components and two connectors.

Distinction between a component's ABT and a connector's ABT is just that a component's ABT is an atomic ABT whose internal structure is unknown, whereas a connector's ABT is known to be an ABT that is itself composed out of other ABTs

Component's Abstract Behavioral Type and connector's Abstract Behavioral Type as models for program components and connectors (glue code) is based on exchanging some passive data with their environment through their communication contact ports on interfaces. Since connectors' ABT is between program components which connect and therefore it must contain some suitable inter-component communication protocol.
The inter-component communication protocol represents interaction among component's observable input and output

$$\text{input}_i \xrightarrow{\text{interaction}_k} \text{output}_j,$$

support only two operations [3], where :

- *sending message* is just write operation and
- *receiving message* is just read operation

and can be yielded by relation and denoted :

$$R_k(input_i; output_j). \tag{2}$$

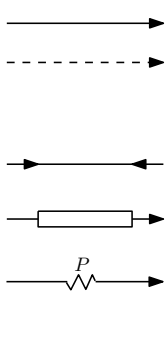
A. Connectors

> *Connector* can be in general defined as a medium of communication with exactly two ends, and a relation that defines its inter-communication protocol through these ends.

We recognize two types of connectors ends:

- *source end* - through this end timed data stream enter into channel;
- *sink end* - through this end timed data stream come out of the channel;

Connector's simplest types are channels (Fig.:2). Channels as a set of simplest *glue code* was defined in [4], [5], [6]. There is no requirement in channels to have defined a source and a sink ends. This idea is satisfies in the case of channel that has two sources or two sink, but there have to be defined behavior of the channel through relation which defines inter-communication protocol of the channel.



Synchronous Channel $\langle \alpha, a\rangle \text{Sync} \langle \beta, b\rangle \equiv \alpha = \beta \wedge a = b$

LossySync Channel
$$\langle \alpha, a\rangle \text{LossySync} \langle \beta, b\rangle \equiv \begin{cases} \langle \alpha, a\rangle \text{LossySync} \langle \beta, a(0).b'\rangle & \text{for } a(0) > b(0), \\ \beta(0) = \alpha(0) \wedge \langle \alpha', a'\rangle \text{LossySync} \langle \beta', b'\rangle & \text{for } a(0) = b(0), \\ \langle \alpha', a'\rangle \text{LossySync} \langle \beta, b\rangle & \text{otherwise.} \end{cases}$$

SyncDrain Channel $\langle \alpha, a\rangle \text{SyncDrain} \langle \beta, b\rangle \equiv a = b$

Asynchronous FIFO_k Channel $\langle \alpha, a\rangle \ \text{FIFO}_k \ \langle \beta, b\rangle \quad \equiv \quad \alpha = \beta \wedge a < b < a^{(k)}, \quad \text{where} \quad b(i) < a(i+k).$

Filter Channel
$$\langle \alpha, a\rangle \text{FILTER}(P) \langle \beta, b\rangle \equiv \begin{cases} \beta(0) = \alpha(0) \wedge b(0) = a(0) \wedge \langle \alpha', a'\rangle \text{FILTER}(P) \langle \beta', b'\rangle & \text{if } \alpha(0) \ni P, \\ \langle \alpha', a'\rangle \text{FILTER}(P) \langle \beta, b\rangle & \text{otherwise.} \end{cases}$$

Fig. 2 Inter-communication protocols of the channels defined by relations.

Synchronous channel: Synchronous channel represents the behavior of any entity in which :

- output data stream is identical to its input data stream $(\alpha = \beta)$;
- element on the output is produced at the same time as element on its input is consumed $(a = b)$;

LossySync Channel: LossySync represents the behavior of any entity that :

- all write operations on the source end of a LossySync channel are immediately succeed;
- if there is a pending take on its sink end, then the written data item is transferred;
- otherwise, the write operation succeeds, but the written data item is lost;

SyncDrain Channel: SyncDrain represents the behavior of any entity that :

- produces no output data stream, because it has no sink end;
- every data item written to its source ends is lost;
- write operation on the one of the end remains pending until write on other end is performed, in other world operations on both ends are performed at the same time $(a = b)$ and write operations can be blocked if it is necessary, to ensure that they succeed atomically;

Asynchronous channel: Asynchronous channel represents the behavior of any entity in which :

- output data stream is identical to its input data stream $(\alpha = \beta)$;
- every element on its output some time after its respective input element is observed $(a < b)$ but before its k^{th}-next input element is observed $b < a(k)$ which means $b(i) < a(i+k)$, for $i > 0$;

Filter channel: FILTER(P) channel represents the behavior of any entity in which :

- only those data items pass through channel that match the pattern P;
- all others data items are always accepted by its source, but are immediately lost;

The infix operator $\alpha(0) \ni P$ denotes whether or not the data item $\alpha(0)$ matches with the pattern P and therefore if $\alpha(0)$ passes through or not. When $\alpha(0)$ does not pass through channel, it is lost, and the ABT proceeds with the rest of its timed data streams.

B. Category of program components

An interface of program component $INF_{pk_{s_q}}$ is a set of all program component contact ports $cp_{pk_{s_{qr}}}$ and can be denoted as :

$$INF_{pk_{sq}} = \{cp_{pk_{s_{q_1}}}, cp_{pk_{s_{q_2}}}, \ldots, cp_{pk_{s_{q_r}}}\} \tag{3}$$

for $1 \leq s \leq n$, $0 \leq q \leq k$ and $1 \leq r \leq m$, while one program component can have several interfaces what we will be distinguish by index q.

All program components contact ports $cp_{pk_{s_{q_r}}}$ and relation among observable input and output timed data streams $R_k(\langle \alpha_i, a_i\rangle; \langle \beta_j, b_j\rangle)$ forms the category $\mathcal{P}$komp, category of program components $\mathcal{P}$komp.

Since program components pk_s communicates with their environment through the contacts ports sending untargeted passive messages, which only supports the operations of writing and reading, we can also say that each program component contact port $cp_{pk_{s_{q_r}}}$ is a type :

$$cp_{pk_{s_{q_r}}} = \langle \alpha(i) : \text{Stream}[T], a(i) : \text{E}\rangle \tag{4}$$

where Stream[T] is a type of the data item $\alpha(i)$ and E is a type of a time moment $a(i)$, for $0 < i \leq n$.

The category $\mathcal{P}$komp, Fig. 3 is a category, where :

- objects are individual component interfaces $INF_{pk_{s_q}}$ which are products of the contact ports $cp_{pk_{s_{q_r}}}$:

$$INF_{pk_{s_q}} = \{cp_{pk_{s_{q_1}}} \times \ldots \times cp_{pk_{s_{q_r}}}\} \tag{5}$$

- morphisms are interactions among observable input timed data stream and output timed data stream in the form :

$$\langle \alpha_i, a_i \rangle \xrightarrow{\text{interaction}_k} \langle \beta_j, b_j \rangle,$$

  which is yielded by relations and can be denoted in the form :

$$R_k(\langle \alpha_i, a_i \rangle; \langle \beta_j, b_j \rangle), \tag{6}$$

  where input timed data stream is a domain of the morphism yielded by relation R_k and output timed data stream is a codomain of the morphism.
- identity can be written as $id_{pk_{s_q}} : INF_{pk_{s_q}} \rightarrow INF_{pk_{s_q}}$,
- composition of the morphisms is defined by the rule :

$$\frac{INF_{pk_{1_q}} \overset{R_1}{\rightarrow} INF_{pk_{2_q}} \qquad INF_{pk_{3_q}} \overset{R_2}{\rightarrow} INF_{pk_{3_q}}}{INF_{pk_{1_q}} \overset{R_2 \circ R_1}{\rightarrow} INF_{pk_{3_q}}}, \tag{7}$$

- composition of morphisms is associative, i.e. for morphisms $INF_{pk_{1_q}} \overset{R_1}{\rightarrow} INF_{pk_{2_q}}$, $INF_{pk_{2_q}} \overset{R_2}{\rightarrow} INF_{pk_{3_q}}$ and $INF_{pk_{3_q}} \overset{R_3}{\rightarrow} INF_{pk_{4_q}}$ holds :

$$R_1 \circ (R_2 \circ R_3) = (R_1 \circ R_2) \circ R_3. \tag{8}$$

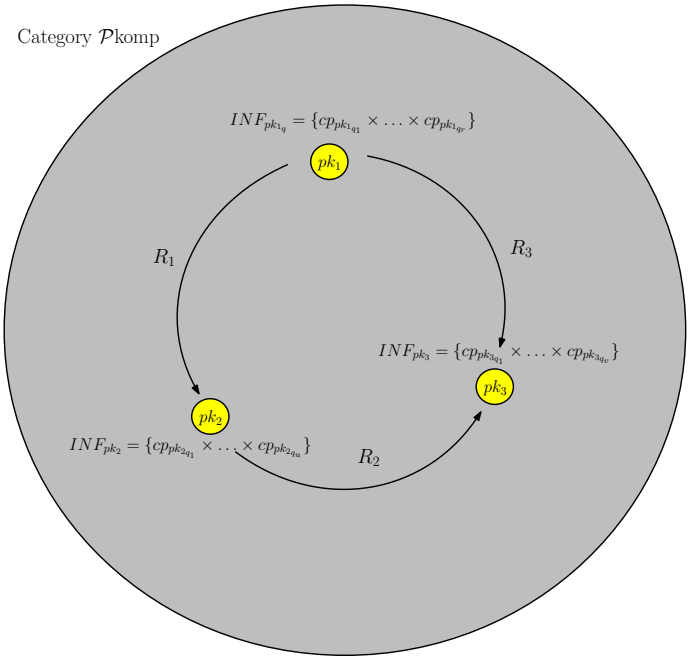


Fig. 3 Category of program components $\mathcal{P}$komp

III. Conclusion

In this paper we have defined category of program components with its properties. Objects of this category are individual component interfaces which are products of the program component contact ports and morphisms are interactions among observable input timed data stream and output

timed data stream which is yielded by relations. Thus defined category form the base model for composition of the program component.

Component composition is not as trivial as it in first seems because instances of components are primitive buildings blocks [7], [8], [9] and all decisions about which components will be used and together composed to build new program application must be made from outside what in generally means that a component cannot internally decide about the communication with another components.

Each connector which glue two program component have to contain inter-component communication protocol. The inter-component communication protocol represents interaction among component's observable input and output

$$\text{input}_i \xrightarrow{\text{interaction}_k} \text{output}_j,$$

support only two operations [3], where :

- *sending message* is just write operation and
- *receiving message* is just read operation

and can be yielded by relation and denoted :

$$R_k(input_i; output_j). \tag{9}$$

In our next research we would like to construct coalgebraic behavioral model for component-based complex program systems.

ACKNOWLEDGMENT

This work is the result of the project implementation: Center of Information and Communication Technologies for Knowledge Systems (ITMS project code: 26220120030) supported by the Research & Development Operational Program funded by the ERDF.

REFERENCES

[1] X. CHEN, H. JIFENG, Z. LIU, and N. ZHAN, "A Model of Component-Based Programming." report No. 350. 2006. UNU-IIST, P.O.Box 3058, Macao.

[2] C. SZYPERSKI, D. GRUNTS, and S. MURER, *Component Software : Beyond Object-Oriented Programming - Second Edition*. Boston, USA: Addison-Wesley / ACM Press, June 2011.

[3] B. KRISTENSEN and D. MAY, "Component Composition and Interaction." in *Proceedings of International Conference on Technology of Object-Oriented Languages and Systems.* TOOLS PACIFIC'96, 1996.

[4] F. ARBAB, "Composition by Interaction." leiden Institute of Advanced Computer Science, Faculty of Mathematics & Natural Sciences, Leiden University.

[5] ——, "Abstract Behavior Types : A foundation model for components and their composition," *Science of Computer Programing*, vol. 55, pp. 3–52, 2005.

[6] ——, "REO: a channel-based coordination model for component compostition." *Mathematical Structures in Computer Science*, vol. 14, pp. 329–366, 2004.

[7] M. JENČIK and D. MIHÁLYI, "Program Components & Abstract Behavioral Types," *Acta Electrotechnica et Informatica*, 2012.

[8] ——, "Formal description of behaviour for large program system," in *CSE 2010 : proceedings of International Scientific conference on Computer Science and Engineering*. Department of Computers and Informatics FEEI TU of Košice, Slovak Republic, 2010, pp. 15–22.

[9] M. JENČIK and V. NOVITZKÁ, "Formal Description of Behaviour for Object-based Programs," in *Electrical Engineering and Informatics 2 : Proceeding of Faculty of Electrical Engineering and Informatics of the Technical University of Košice*. Faculty of Electrical Engineering and Informatics, Technical University of Košice, Slovak Republic, March 2011, pp. 223–228.

Visualization of PCB Fabrication in Education

[1]*Ondrej KAINZ,* [2]*Slavomír KARDOŠ*

[1] Department of Computers and Informatics, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Slovak Republic
[2] Department of Technologies in Electronics, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Slovak Republic

[1]ondrej.kainz@student.tuke.sk, [2]slavomir.kardos@tuke.sk

***Abstract* — The contribution describes visualization approach in educational process of electrotechnology, aimed to PCB fabrication. The Adobe Flash tool is described as the tool for particular animations development. The latest knowledge in the field of animation technology was used during development. The animations of technological procedures are the "opus locatum" as the support of study text and video materials in the branch of electrotechnology.**

***Keywords* — PCB fabrication, Adobe Flash, e-learning, virtual laboratory**

I. INTRODUCTION

In recent times the level of computer animation and graphics allows to convert almost any idea into the visual form. In the learning process is often necessary to demonstrate theoretical knowledge practically.

Education in general can be considered as complex interaction process between educator and target group or learner. During this process it may be appropriate, sometimes even necessary to dynamically demonstrate the problem. Solution for this can be provided using computer animations and thus to provide the dynamic learning experience for learners. Animations can significantly enhance learning, motivation, attitudes and also reduce time needed for learning or gain attention of learners. Animations can show relationships between objects and ideas, simulate the results of actions, simulation, show sequential steps in procedural task, explain difficult concept and make the abstract concept as concrete. Recent computer technology now allows animation to offer a high degree of learner control [1][2].

A computer animation is very complex area of computer graphics. For the animated components we consider lines, forms, shapes, objects and its parameters in basic raster or vector form capable of animation in 2D or 3D space for a period of time using computers. These components are used in computer animations. Computer animation is any computer-based computation used in producing of images, intended to create the perception of motion.

Flash animations are timeline based vector oriented animations with interactive behavior, stored in SWF file format [3]. Multimedia development environment, Adobe Flash, is based on vector graphics using frame-based model. Combining text, all images formats input support, audio, animation and multiple video formats support, the tool can be used to create animations, to display video on a website, to create handheld applications or build a complete web-based application. Due to the adjustable compression support the file size can be adapted. Animations in Flash can be created frame by frame or using motion, shape and classic tweening. In Flash can be created and rendered basic 3D animations or imported from 3D modeling software such as 3ds Max as well. Animations may also be nested within the object. Embedded object-oriented language - ActionScript can be used to control the animations and components contained within. It has webcam and microphone full support [4][5][6].

II. ANIMATION COMPONENTS APPLICATION IN LEARNING PROCESS OF ELECTROTECHNOLOGY

Theoretical knowledge is possible to describe using animations also visually and by this way to emphasize and highlight the principle of operation and explain the given problem. Moreover, it is possible to achieve the learning process for target groups more interesting and attractive.

One advantage is the ability to visually demonstrate the problem, which ultimately contributes to a better understanding of problems and facilitate the teaching-learning process. Animation, in our case Flash animation, can be used not only in learning process itself, but also on websites, such as a virtual laboratory.

Animations in learning process of electrotechnology have wide range of use. Typically in the educational process is necessary to present the given problem theoretically, but practically as well. Practical implementation of a given problem may be complex and very expensive, or require other means, which occupy a longer period of time. For example, for a demonstration of reflow soldering process is required reflow oven, such equipment can cost several thousand euros and have high power consumption. Another possibility to demonstrate mentioned process is to use the external sources, for example in companies that use this technology. Both mentioned kinds of the implementation are either expensive or time consuming and may complicate the teaching process.

Flash animations allow easily and effectively present given problems without owning of expensive equipment, they can be implemented simultaneously with the theoretical interpretation or used in virtual laboratory available on the web, in this way can be animations viewed almost everywhere where internet connection is available.

Animations are therefore highly effective way to improve the teaching process and make the explanation of theoretical part easier for the lecturers and also for the learners. Using of animations it may also be increased the attention of the target group or learners and learning process may become more attractive.

III. Design of PCB fabrication animated components

Development of animations, starting from the analysis through the design and implementation itself is described in this chapter. Storyboard layout, object definition and key-frame specifications was analyzed considering the goals of animations, then animation features was described in design part and finally implemented.

Animations should describe the given technical procedures of electrotechnology correctly and clearly. Concerning the fact that they will be used during lectures or in virtual laboratory, the resolution should not be higher than 800x600. Size of animation to be 700 (width) x 500 (height) pixels which provides enough space for presentation was chosen for all animations. Higher resolution may have negative influence on computer resources.

Duration of animations is longer, consisting of more parts. These parts should be available in menu. Appropriate would be also possibility to speed up or down playback, stop it or pause it. Control of playback should be easy, clear and intuitive.

Detail creation is described in example animation of surface and through-hole components placement and wave soldering. Before wave soldering, the surface mount components are glued to the bottom with adhesive, then the adhesive is cured and printed circuit board (PCB) is rotated upside-down in order to place thought-hole components. During soldering a wave of molten solder is put into contact with the PCB which is moving on conveyor [7][8].

The visible animation area, stage, is divided into Front view and Top view and menu. Side view is also used for some animations. Last part of stage was left for menu as a control the playback.

In the next were drawn the main parts, i.e. PCB in Front view and PCB in Top view, axis, Front view background, Top view background and PCB holders. Almost all objects were always in 2D representation, although PCB in Top view and holders were during upside-down rotation using perspective view, this was resolved using Adobe Flash 3D tools.

Other objects that will occur later in the animation are adhesive dispenser, component placer, adhesive, capacitor, resistor, adhesive oven, PCB holder for 3D, through-hole resistor, thought-hole capacitor, molten solder wave. Individual object was PCB during 3D rotation in Top and Front view; components and PCB were nested in the movie clip in order to rotate PCB.

Component molten solder wave was created as nested animation, literally animation in animation, feature provided by Adobe Flash. Except axis and PCB holders, all objects were drawn in Adobe Illustrator, vector graphic editor, which provides better drawing tools and have good import options from Illustrator into Flash. All shown objects were drawn according to model of real devices to bring the most objective visualization of process.

Key-frame specifications were quite complicated due to large number of components present in animation. Movement in “Front view” had to be synchronous with movement in “Top view”. Whole animation has 6470 frames; at a rate of 24 frames per second it is almost 270 seconds. Playback speed might differ regarding machine resources.

For an adhesive injection, the shape tween technique was used, meaning that shape had to be drawn in key position and Adobe Flash fill the space with in-betweens, similarly as in motion tween. To simulate movement along Y axis, the scaling of an object was used, during animation playback can be seen that object is approaching or moving away. For movement along X axis and Z axis, X and Y coordinates were used. Alpha transparency of layer with aim to simulate heating process was used during heat adhesive curing.

Moderately complex part was during modeling of three dimensional rotation of PCB in "Top view". In the movie clip had to be nested layers representing sides of PCB and also sides of components and PCB holder.

Camera view was switched before soldering itself to "Solder wave view" to better illustrate the wave soldering. In this way was possibility to observe the deposition of molten solder on the components. Soldering wave contains nested animation.

After finishing the animation part, the control button Playback speed potentiometer was designed using object oriented language ActionScript 3.0. In Fig. 1, the initial animation stage is shown and Fig. 2 shows end of the animation with components soldered to the PCB using wave soldering.

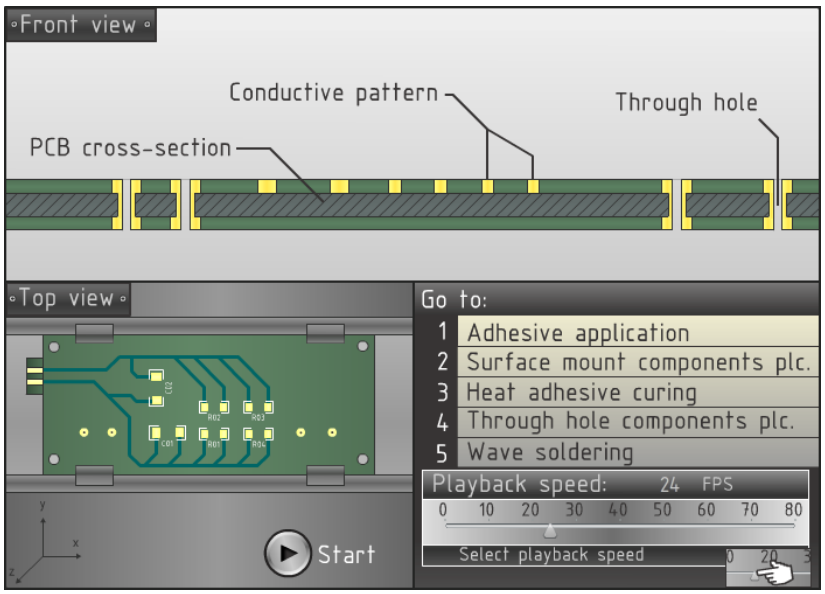


Fig. 1 Initial stage of the animation.

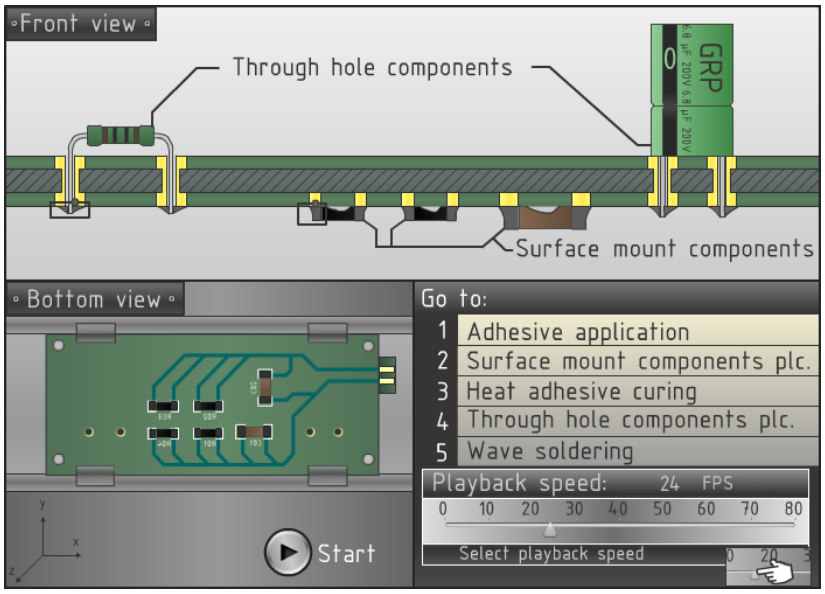


Fig. 2 The last step of the animation.

To make control of playback speed be possible a slider with simple "if" statement was created. There is a notable difference between these playback speeds. On computers with minor

resources would be proper to use 24 or 50 FPS instead of 80 FPS.

In total were created four animations, describing four different techniques. The first and the second animation have common components; both represent two different types of soldering techniques. As well the third and fourth have common components, these two represent thick-film masking techniques.

Second animation, wave soldering, was described into detail. In this part will be briefly described first animation – reflow soldering, third animation – photolysis and fourth animation – screen printing.

Reflow soldering technique is used mainly for soldering of surface mount components. Reflow soldering sequences provide maximal mounting density without creation of any solder bridges. Before reflow soldering has to be applied solder paste on joints where components are then placed. After this process PCB with components is conveyed to the reflow oven, the solder paste is activated by the temperatures reaching up to 75-80°C. Next the paste is soldered using infrared radiation, laser, heated hot-air or pulse soldering using a resistance-heated strip or using U-strap support [7]. This animation (Fig. 3) was created in the similar manner as the above one described. There was necessary to animate three steps. Objects, such as PCB, except conductive connections, component placer, were used the same as in the wave soldering animation, as well windows arrangement was the same. Actually this animation predeceased wave soldering animation, this was created as the first one. Movement is realized again mainly via motion tween technique, due to less complexity key-frame table was not developed.

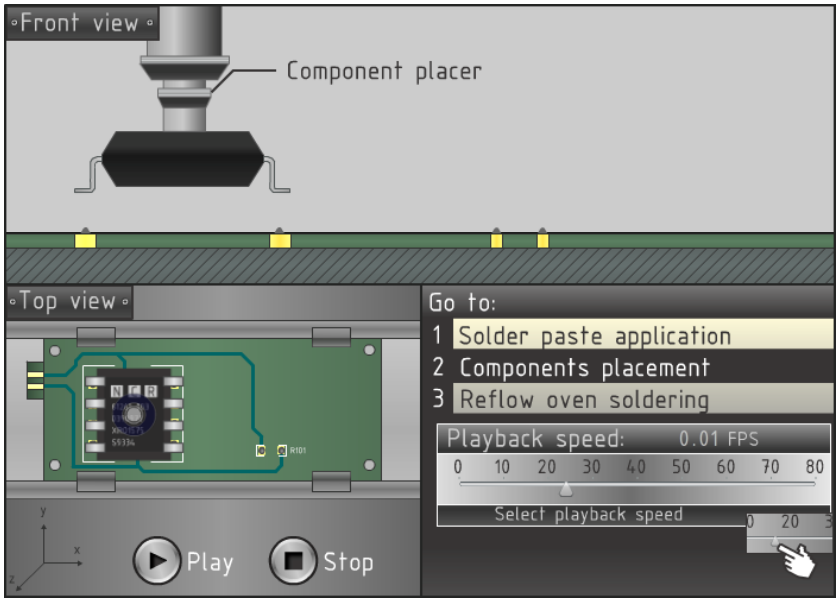


Fig. 3 Reflow soldering animation in progress.

Photolysis belongs to the subtractive procedure of circuit board manufacturing; it is classical approach where redundant copper is etched from board and conductive pattern is created. In production, on board is laminated or spayed the layer of photoresist, on one side or both sides. Board has to be fixed, for example using vacuum fixation, positive or negative mask is applied and then photoresist layer is activated by ultraviolet radiation exposure and polymerization process takes place. This causes photoresist layer to be washed away, leaving photoresist only on places covered by positive or negative mask. This process is called developing, after that comes etching which washes away the copper layer. The last stages are cleaning and removal of photoresist from conductive pattern [7]. On Fig. 4, sample from Photolysis animation is displayed, where copper etching is in progress.

IV. Conclusion

The work refers about implementation of computer animations in electrotechnology education practice. The utilization of animation is used from simple blinking button to the simulation of physical process. Possibility to visualize given problem may make educational process easier and more interesting for target group, learners and lecturer as well. Using animation instead of real devices may also save resources of given institution, in this case of the university. The Adobe Flash software was chosen, which provides best options to create

suitable animations. Final animations are playable on almost all machines, since Flash is wide spread technology. Created were four animations, each describing different technological procedure. Methodology for creating of animations was used parts of the Systems Development Life Cycle and Animation Design. As a model animation was chosen "Wave soldering", since it contained majority of objects and techniques used. During animation design stage there was effort to create the animation in accordance with real devices. These animations are used as the learning aids for improvement of educational process and may be used in virtual laboratory, what is in the stage of base enlarging.



Fig. 4 Photolysis animation in progress.

ACKNOWLEDGMENT

This paper was developed with support of the project „Vývoj unikátneho nízkoenergetického statického zdroja pre elektrosystémy", ITMS 26220220029, that is co-financed from Structural Funds EU ERDF within Operational programme Research and Development V-2008/2.2/01-SORO and preferred axis 2 Support of Research and Development.

REFERENCES

[1] FENRICH, Peter. Creating Instructional Multimedia Solutions: Practical Guidelines for the Real World. Santa Rosa: Informing Science Press, 2005. pp. 154-157. ISBN 83-922337-1-9.
[2] LOWE, Richard. Interrogation of a dynamic visualization during. Learning and Instruction. [Online] [Cited: December 3, 2010.] http://www.diliaranasirova.com/blog/PSYC579/pdfs/04.2-Lowe.pdf.
[3] HAMLIN, J. Scott. Grafika, animace, kouzla na Webu. Brno: UNIS Publishing s.r.o., 2000. pp. 26-34. ISBN 80-86097-45-5.
[4] GROVER, Chris. Flash CS5: The Missing Manual. Sebastopol: O'Reilly Media, 2010. pp. 13; 18; 20-22. ISBN 978-1-449-38025-0.
[5] CHUN, Russell. Adobe Flash Professional CS5 Classroom in a Book. San Jose: Adobe Press, 2010. pp. 1; 11-16; 21-29. ISBN 978-0-321-70180-0.
[6] SHUPE, Rick and ROSSER, Zevan. Learning ActionScript 3.0: A Beginner's Guide. Sebastopol: O'Reilly Media, 2008. pp. 4-5. ISBN 978-0-596-52787-7.
[7] SLOSARČÍK, Stanislav, PIETRIKOVÁ, Alena and BANSKÝ, Juraj. Základy technológií v elektronike. Košice: Vienala, s.r.o., 2006. pp. 19-25; 57-58; 64-68; 89. ISBN 80-89232-08-6.
[8] PRASAD, Ray P. Surface mount technology: principles and practice. New York: Springer-Verlag, 1997. p. 535. ISBN 978-0-412-12921-3.
[9] BECKERMAN, Howard. Animation: the whole story. New York: Allworth Press, 2003. pp. 4-7. ISBN 978-1-58115-301-9.
[10] PARENT, Rick. Computer Animation Complete. Burlington: s.n., 2009. pp. 4; 27-35; 278-285. ISBN 978-0-12-375078-5.
[11] Intro to Animation Techniques - A Look at the Three Main Types of Animation. [Online] [Cited: April 12, 2011.] http://movies.about.com/od/animatedmovies/a/animated-techniques.htm.
[12] LIEU, Dennis. Visualization, Modeling, and Graphics for Engineering Design. New York: Delmar, Cengage Learning, 2009. pp. 19-3. ISBN 978-1-4018-4249-9.
[13] SMITH, Ian F.C. Intelligent Computing in Engineering and Architecture. Heildelberg: Springer-Verlag, 2006. pp. 258-260. ISBN 978-3-540-46246-0.

Recording of Information on VoIP Calls to Calendar System at DCI

Katarína KLEINOVÁ, Peter FECIĽAK, Michal BENČKO

Department of Computers and Informatics, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Slovak Republic

katarina.kleinova@tuke.sk, peter.fecilak@tuke.sk, michal.bencko@gmail.com

Abstract **– A logging of the performed VoIP calls is an important part of the VoIP telephony. Nowadays there are more and more equipment, tools and applications, with which it is possible to record details about performed VoIP calls for later analyzing if necessary. The main goal of this paper is to describe an extension of VoIP analyzer application for an ability of recording details about performed calls to the Google calendar. This paper deals with a detailed analysis of the existing VoIP analyzer application with the mechanisms for accessing Google calendar and consequently by the extension of the application of the necessary functions and methods. Another objective of this paper is to present the recorded data about VoIP calls on a website. For this purpose web application was created, that displays details about the performed VoIP calls and with each refreshing of the website its content will be updated.**

Keywords **– Google Calendar, logging a VoIP calls, network traffic, SIP, VoIP**

I. Introduction

With the launch of the Internet and the discovery of opportunities for transmitting the voice over IP protocol, development of IP telephony went considerably further. The transmission of voice using Internet Protocol (Voice over Internet Protocol), also called VoIP, IP Telephony, Internet telephony is the transmission of human voice communication through the Internet or another network based on IP protocol. An important part of VoIP communication is also the logging of the performed calls. The main goal of recording the VoIP calls is the possibility to analyze the recorded information in the future when it is needed. Most common tools and applications record information like source/destination addresses, beginning/end of the VoIP call, call duration and so on.

The extension of VoIP analyzer application should be able to identify the beginning and end of a call and based on these data calculate the length of the performed VoIP call [1].

II. Capturing of VoIP communication

Sending data in an IP network is based on identifying the source and target devices. The process of capturing VoIP calls via additional application, which is not participating in the communication, is called packet sniffing. Transfered data from the sender passes through various connecting elements to the recipient. When considering common circumstances, the third device, which can be for example computer (does not participate in the data transfer) should not get packets containing valuable information.

However this option exists in the following cases:

- Use of inadequate topology sharing internet connection through another computer.
- Use of improper connecting elements (e.g. multiple-port repeaters).
- The desired effect for the purpose of measurement (connecting elements are configured to allow capture of communication).

If it is a wished effect, these confidential information can also get only on the network card of the target device that would like to eavesdrop VoIP communication. In most cases this feature is not activated yet on the network card. Information of this type is discarded, because its destination address is different than that which is assigned to the computer. The network card detects this type of information and discards the packets. Interception of communications is therefore possible either on the device through which information is transferred (via internet connection sharing) or as in case of VoIP analyzer on PC, to which the information of this kind should not ever

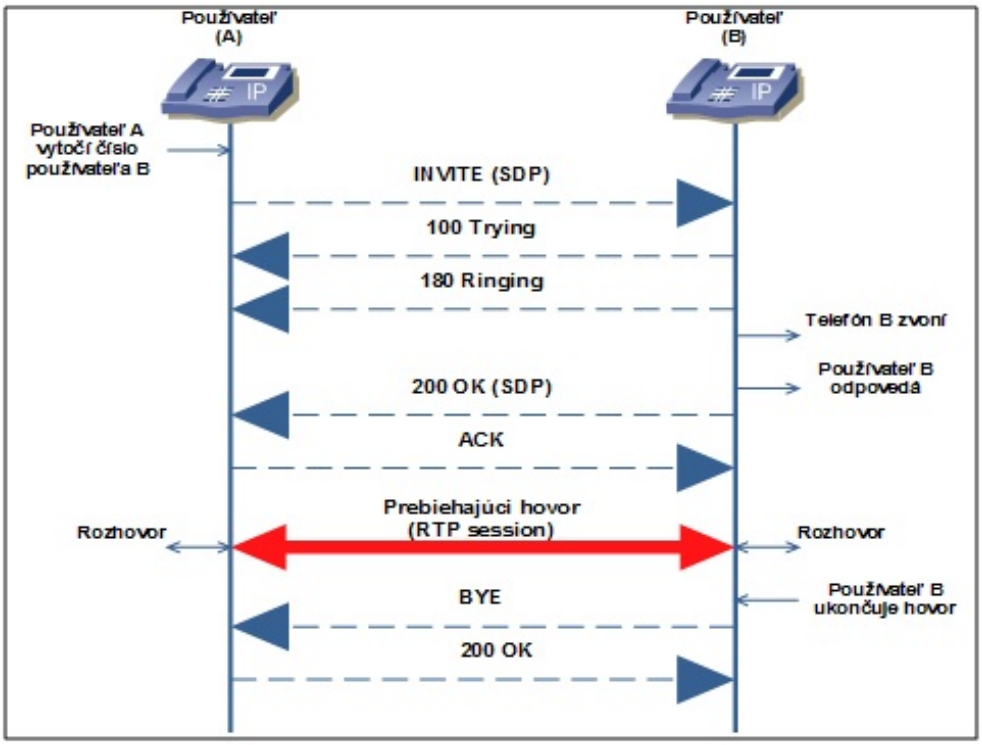


Fig. 1 Direct SIP call flow between two IP phones.

get. In the second case, when the packet flow do not pass through our network interfaces it is important to turn off hardware filter of the network card to make packets capturing possible. It is done, because the hardware filter is set to discard all packets, which are destined for the network interface. After switching off the hardware filter, network card gets into a state called promiscuous mode. It is the state of the network card, when all packets are delivered for processing in the operating system, regardless from where and where they were send. The process of the packet capturing is based on ignoring the destination port stored in the packet header. In this way, every packet that arrives at the network interface card is read by the tracking application (wireshark, tcpdump, etc) [2]. Thus all unencrypted communication such as important transferred data (i.e. source address, destination address, used protocol, in case of website its address and content) is possible to read. From forms filled out on the websites it is possible to get all the entered data. In case of VoIP analyzer, attention is focused on the header of SIP and RTP protocols. The Fig.1 shows the SIP messages that are exchanged between two SIP-enabled IP phone devices during a basic call establishment [3], [4].

- **INVITE** - Basically, the SIP call establishment is simple, it starts with a SIP INVITE message, which is sent from the calling party (in the above example user A) to the called party (user B). The message invites the called party to participate in a SIP session.
- **100 Trying** - The called party replies with a 100 Trying message indicating that the INVITE message was received and now some kind of action is being taken to process it. The 100 Trying message does not indicate that the user has been located.
- **180 Ringing** - If the invited user was located, the called party performs two operations: it applies a ringing signal to the phone and, in parallel, sends a 180 response to the calling party informing it that phone B is now ringing. Upon receiving the 180 response, the calling party is aware that phone B is ringing and therefore plays a ring-back tone to phone A.
- **200 OK** - When user B answers the call, the called party sends a 200 OK response informing the calling party that the call was accepted.
- **ACK** - The calling party acknowledges that the called party has answered by issuing an ACK message.
- At this point data is exchanged between these parties. The protocol that handles data transfer is the RTP Protocol.
- **BYE** - Finally one of the parties (in the above example it is user B) hangs up the phone, which causes a BYE message to be sent.
- **200 OK** - The other party receives the BYE message and sends 200 OK to confirm receipt

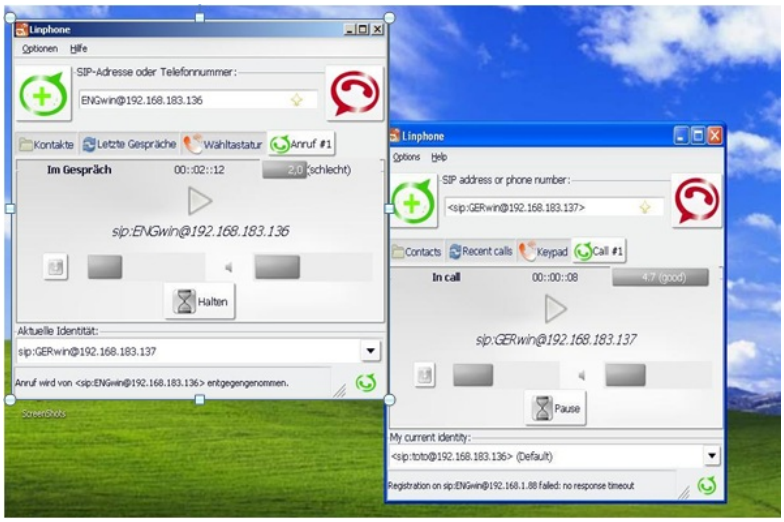


Fig. 2 Communication between two softphones Linphone in virtual OS Windows XP.

of the message. At this point the call is over.

Before the first SIP message is sent, a TCP handshake is established between two endpoints. Also note that every SIP message is acknowledge by a TCP ACK packet. These reliable transfer mechanisms are an integrated part of the TCP transport layer. On the other hand, when SIP is used over UDP, the SIP application must handle the retransmission. The number of SIP messages is much less than the number of RTP messages. This, of course, is logical since RTP is used to transfer voice over IP, and typically every RTP packet carries around 20 - 40 msec of voice samples.

III. Structure of the VoIP analyzer

Voice over IP analyzer was for the proper functioning of the program designed like an application that provides a number of mutually independent operations simultaneously [5]. It is a multithreaded application and has the following features:

- Identification of the start and the end of the call. Based on these two details it is possible to determine call duration.
- Identification of the calling and the called person.
- Recording of this information to the Google calendar and their interpretation on the website.

To implement the changes on VoIP analzyer development environment Microsoft Visual Studio 2010 was used, which was installed in Windows 7. The operating system also serves as a virtual guest system solutions for two operating systems Windows XP which are running under VMWare application. Both virtual machines were used to simulate a simple network environment with a small, theoretically no packet loss rates. As the telephone equipment software solution of IP phones ”Linphone” were used that do not need to run a SIP server and can directly communicate with each other. These software devices were installed on both virtual operating systems Windows XP. For monitoring all network traffic and communication between the phones application Wireshark was used. On the Fig. 2 there are presented two software phones Linphone while communicating to eachother in virtual Windows XP.
After initiating the VoIP call the VoIP analyzer sends the request for the initial ”info” packet to the server and at the same time there is presented a list of available network cards. After selecting the network interface from the presented list and opening communication between software IP phones, VoIP analyzer captures INVITE packet. After capturing INVITE packet analyzer expects OK and ACK packets. When the ACK packet is received the VoIP analyzer stars to monitoring network traffic, which is flowing through the selected network interface and simultaneously measuring and evaluating the parameters of the network traffic. These values are measured and presented in real-time. On the Fig. 3 running VoIP analyzer measuring network traffic is displayed.

When the call is interrupted, VoIP analyzer captures BYE packet and starts the application for importing data into Google calendar. After the successful creation of new event in the calendar, which consists of information about last performed VoIP call message about successful importing data into the calendar could be viewed in console window for 5 seconds (Fig. 4).

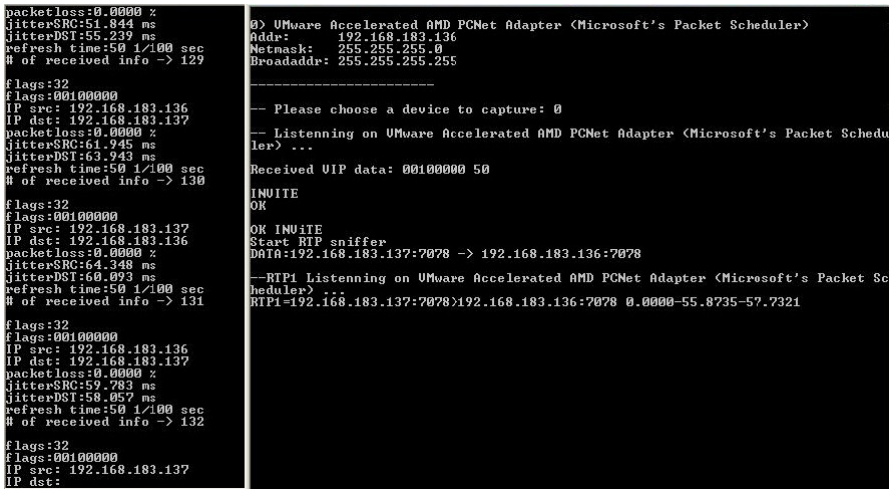


Fig. 3 VoIP analyzer monitors network traffic on selected network card.



Fig. 4 Successful importing of the information about performed VoIP call into the Google calendar.

For the illustration of storing information about performed VoIP calls was created testing Google account "voipanalyzer". Application creates events for every single captured VoIP calls (Fig. 5). These events include following information:

- Indication who was calling and who called person (SIP address).
- Details of call time (beginning and end of the call).
- Detail about call duration.

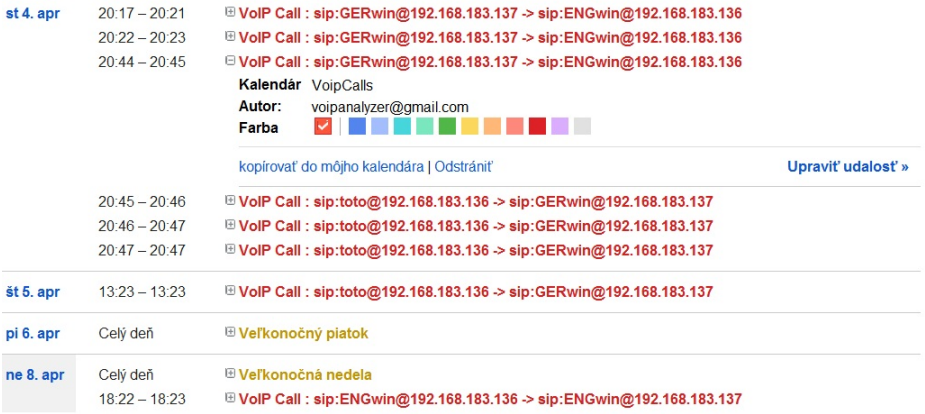
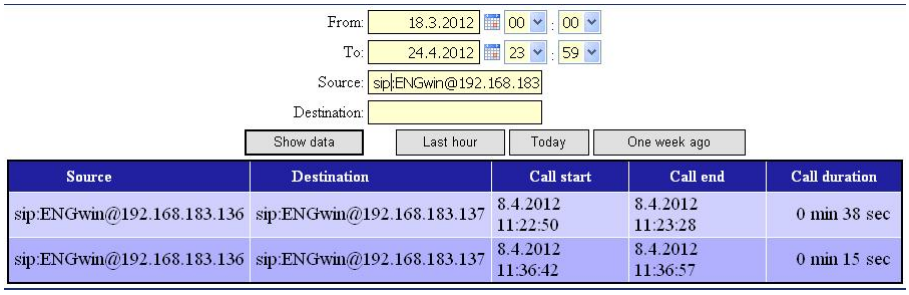


Fig. 5 Created events about performed VoIP calls.

As it was stated previously this information should be presented also on website. For that reason web application was created which is able to present stored information about performed VoIP calls on website. VoIP analyzer saves this information in log text files from which this application can read them. Content of the website is updated with every single refresh (Fig. 6). Application allows user to filter traffic based on time scale (from-to) or based on source or destination SIP address. It also allows to view calls made in last hour, last week and calls made during the day when the data is viewed. In the next picture is possible to see this website with the content.



Source	Destination	Call start	Call end	Call duration
sip:ENGwin@192.168.183.136	sip:ENGwin@192.168.183.137	8.4.2012 11:22:50	8.4.2012 11:23:28	0 min 38 sec
sip:ENGwin@192.168.183.136	sip:ENGwin@192.168.183.137	8.4.2012 11:36:42	8.4.2012 11:36:57	0 min 15 sec

Fig. 6 Details about performed VoIP calls presented on website.

IV. Conclusion

The main goal of the paper was modification of VoIP analyzer application of recording information about performed VoIP calls. All objectives were managed to meet, resulting in extended VoIP analyzer application, which is capable to record data about participants of the call. The newly created application GCalendarEntry is used for entering the data about performed calls to the Google calendar by creating new events. VoipWeb web application was created based on the requirement to present the recorded data about VoIP calls on the website. Despite some differences in information processing in operation systems Windows and Linux, VoIP analyzer is fully functional application on both platforms, without the need of recompilation. Modified VoIP analyzer in cooperation with applications GCalendarEntry and VoipWeb is now able to measure the quality of services provided in VoIP communications and stores data about VoIP calls to Google calendar and present them on the website.

V. Acknowledgment

This work is the result of the project implementation: Development of the Center of Information and Communication Technologies for Knowledge Systems (ITMS project code: 26220120030) supported by the Research & Development Operational Program funded by the ERDF.

References

[1] M. Tkáčik, K. Kleinová, J. Janitor, and P. Feciľak, "Controlling of networks operation by utilizing an applications measuring within the frame of VoIP protocols," *7th Int. Conference on Emerging eLearning Technologies and Applications, ICETA 2009 - Conference Proceedings, Stara Lesna*, pp. 137–143, November 2009, ISBN 978-80-8086-127-8.
[2] *TCPDUMP 1997*, Online: http://www.ethereal.com/docs/man-pages/tcpdump.8.html, 04 December 2012.
[3] *Voice over IP - SIP*, Online: http://www.protocols.com/pbook/VoIPFamily.htm#SIP, 30 March 2009.
[4] *SIP call flow 2003*, Online: http://sipandvoip.com/voip-protocols/sip-protocol/sip-call-flow/, 20 March 2012.
[5] M. Tkáčik, *Riadenie sieťovej prevádzky využitím aplikačných meraní v rámci VoIP protokolov, Diplomová Práca*, KPI FEI TUKE, Košice, May 2009.

GPGPU Technology Use for Security Solving in Network Environment

Jozef KOKORUĎA, Branislav SOBOTA

Department of Computers and Informatics, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Slovak Republic

jozef.kokoruda@gmail.com, branislav.sobota@tuke.sk

***Abstract* — Security and data protection is one of the basic problems in computer networks. One of the security measures for data protection is using some encryption method in data transfers. Many of these methods are very computationally demanding so it is important to find a way how to speed up the data encryption process. One of the possible solutions is use of graphics processing units. Graphics cards become very powerful parallel computing devices. The objective of this work is to involve the graphics card to compute some of the encryption method.**

***Keywords* — GPGPU, OpenCL, encryption**

I. INTRODUCTION

Despite the fact that GPGPU is relatively new area of computer science, has undergone rapid development and it has great potential. Graphics cards can be used not only for their primary function, but also for complex and computationally intensive calculations. Solutions such as CUDA and OpenCL together with increasing computing power of graphics cards may cause GPGPU expansion into commercial sector and normal user applications [1]. But parallel GPU architecture is not suitable for all kinds of calculations. GPU-s are suitable for video encoding, mathematical operations, physical simulations [5] and encryption. Data protection is nowadays one of the most important issues for many companies and individuals. Internet and generally computer networks expansion also mean an increased need to protect data from misuse. Encryption is one way of data protection. With rapid development in computing, many encryption methods become ineffective so there is therefore a need to use complex ciphers, what brings increased demands for computing power. The topic of this contribution is to use GPU computing power to compute some of the encryption methods. The goal is to analyze and select appropriate ciphers and design and develop an application that would encrypt data transmitted over the network using GPGPU technology. The final step is to make measurements and comparative study between CPU and GPU using this application.

II. ANALYSIS

The application consisted of three parts:

- Client/server application for transmitting data over network
- Implementation of chosen ciphers on GPU
- Host application for communication with user

To solve client/server part, it was necessary to choose appropriate network programming interface for Microsoft Windows. The Windows sockets API -Winsock - were chosen. Winsock defines interface between Windows TCP/IP client application and the underlying TCP/IP protocol stack. Winsock is supported in all versions of Microsoft Windows and is able to communicate with UNIX network applications which are using UNIX sockets.

In second part it was necessary to choose ciphers which could be effectively computed on GPU. Appropriate part of stream cipher RC4 and block cipher XTEA [2] were chosen for implementation. RC4 is most widely used stream cipher. RC4 consist of the key-scheduling algorithm and pseudo-random key generation which are used for key generation. The last step

is operation XOR between plaintext and key. XOR is used in almost every stream cipher so this operation was implemented on GPU. As a representative of block ciphers, XTEA algorithm was chosen. XTEA belongs to a large group of ciphers based on Feistel network. Because block cipher by itself allows encryption only of a single block of data, mode of operation have to be used for longer plaintexts. ECB mode was chosen - plaintext is divided into blocks and each block is encrypted separately. It allows simple implementation for GPU architecture.

Currently there are three GPGPU solutions available: CUDA, OpenCL and DirectCompute. Because OpenCL is open standard for cross-platform parallel programming, OpenCL were chosen as a solution for GPGPU implementation. OpenCL is currently supported in all major operating systems such as Windows, Linux, MacOS. OpenCL is also supported by all major companies on CPU and GPU market - Intel, AMD, NVIDIA, VIA and many more.

I. Solution and Results

A. Client-server

Winsock requires low-level languages, so the C language was used for creating necessary functions. The mechanism of communication between server and client is described in follow steps:

1. Daemon is started on server side. Daemon listens on the network and awaits the client to connect.
2. Client sends request for connection.
3. Server accepts the request from client and sends acknowledgement.
4. Client sends data to the server. Data are split into appropriate parts before sending.
5. After the last part of data client sends a request to terminate the connection.
6. Server accepts request, sends acknowledgement and terminate the connection.

B. OpenCL

OpenCL requires several steps before the main calculation can be done [3]. It must be ensured that the appropriate device was found in the system and then correctly set to the task. The mechanism of setting the device for calculation is described in follow steps:

1. Initialization of OpenCL devices. It requires several steps:
    a. First step is creating the context to run OpenCL on available devices in system. It is possible to choose GPU or CPU or all devices available in system to make calculations.
    b. Get the list of available devices associated with this context.
    c. Create a command-queue on selected devices.
2. In the next step GPU memory is allocated for source and output vectors. These vectors are initialized from CPU memory.
3. OpenCL allows applications to create a program object using the program source or binary. In this step the source code is loaded into the program object. Then program can be built using OpenCL JIT compilation. The last step is to create a kernel object from compiled OpenCL function.
4. In the next step the GPU memory is associated with Kernel arguments.
5. Kernel is now prepared to be launched on device. It is necessary to specify work-item dimensions to be executed.
6. After the calculation, the output can be copied from device memory back to CPU memory.
7. The last step is to free all memory objects.

OpenCL kernel is function which is executed on OpenCL device. OpenCL kernels are based on C99 language, so the programmer can define own functions like in C/C++. Difference is that the functions are calculated in parallel.

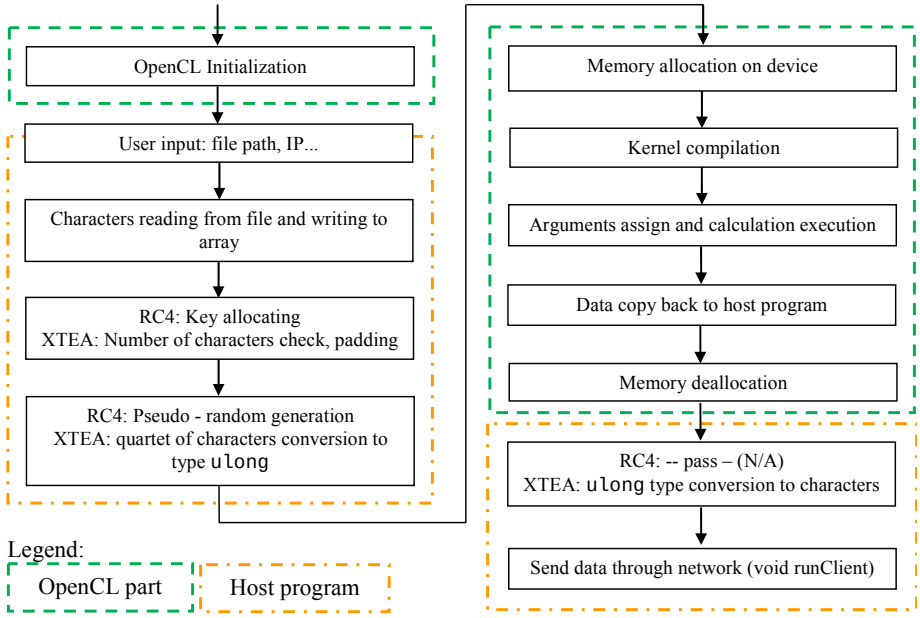


Fig. 1 The program structure with RC4/XTEA encryption implementation

C. *Results*

Both implemented encryptions were tested on three GPUs - integrated NVIDIA GeForce 8200, GeForce 8600GT and GeForce 8800GT - and compared with CPU AMD Athlon X2 4450e 2,3GHz. As the source files were used standard text files with different amount of characters (file1: 92500 characters, file2: 350000 characters, file3: 2100000 characters). The results are depicted in Table 1 and Table 2 and in Fig. 2. None of the graphics cards were able to get better results than CPU. There are several possible reasons for this outcome. Most likely it is caused by incorrect choice of implemented ciphers. Calculation on GPU could not eliminate the time needed for copying data to device and back. Using more computationally demanding ciphers would deceased the impact of copying and GPU could achieve better results.

Table 1
RC4 encryption

	CPU(ms)	GF 8200(ms)	GF8600GT(ms)	GF8800GT(ms)
file1	715	1293	1832	1248
file2	1263	1894	2364	1965
file3	6428	6770	7342	6803

Table 2
XTEA encryption

	CPU(ms)	GF 8200(ms)	GF8600GT(ms)	GF8800GT(ms)
file1	768	1332	1689	1276
file2	1748	1950	2134	1996
file3	6318	N/A	N/A	N/A



Fig. 2 RC4/XTEA encryption for different GPU

II. Conclusion

This paper presents the analysis and selection of appropriate ciphers and design and

application development that would encrypt data transmitted over the network using GPGPU technology and it compares GPU and CPU efficiency in this application. The example of use this application is shown in Fig. 3 and Fig 4. The example of used source text file is shown in Fig. 3 and the examples of this text file encrypted using both implemented encryption methods (RC4 and XTEA) are shown in Fig. 4.

Although using GPGPU to encrypt data did not bring better performance than CPU, as shown in many studies (e.g. AES implementation [4]), combination of GPGPU and encryption has great potential, but requires a suitable choice of ciphers. The question for the future remains the development of new ciphers that could effectively utilize the potential of the GPU, because most currently used ciphers were published in the nineties, when the area of GPGPU was only at the beginning.

Approach of multi-core processors, APUs and ARM processors can promote solutions such as OpenCL (it allow to use different types of processors to perform calculations). Also the approach of cloud systems will require high computing performance to protect and encrypt data, which may be scope for GPGPU application.



Fig. 3 The example of source file



Fig. 4 The examples of file encrypted using RC4 (left) and XTEA (right)

ACKNOWLEDGMENT

This work was supported by the Slovak Research and Development Agency under the contract No. APVV-0008-10: Modelling, simulation and implementation of GPGPU-enabled architectures of high-throughput network security tools.

REFERENCES

[1] Sobota Branislav, Perháč Ján, Szabó Csaba: Distributed GPGPU - new trends; GCCP 2009 : 5th International Workshop on Grid Computing for Complex Problems : October 26-28, 2009, Bratislava, Slovakia. - Bratislava : ÚI SAV, 2009 S. 138-145. - ISBN 978-80-970145-1-3

[2] Needham Roger, Wheeler David: Tea extensions. Technical report, Computer Laboratory, University of Cambridge, October 1997

[3] Khronos OpenCL Working Group: The OpenCL Specification, 2011, 308 p

[4] Gervasi Osvaldo, Russo Diego, Vella Flavio: The AES Implementation Based on OpenCL for Multi/many Core Architecture. In International Conference on Computational Science and Its Applications. 2010. pp. 129-134

[5] Sobota, B., Perháč, J., Straka, M., Szabó, Cs.: Application of parallel, distributed a network computer systems for solution of large graphics data processing; elfa Košice, 2009, ps. 180, ISBN 978-80-8086-103-2 (in slovak)

How to Express an Action Semantics of Programming Languages in Categories

Matúš KOLESÁR, Viliam SLODIČÁK

Department of Computers and Informatics, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Slovak Republic

matus.kolesar.2@student.tuke.sk, viliam.slodicak@tuke.sk

Abstract – **This work deals with the action and categorical semantics of programming languages. It contains a brief theoretical analysis of both methods, which is oriented on their basic properties and main principles of expressing the semantics of programs. The next section focuses on proving of their relationship, which is equivalence. Achieved results are applied in the last section of the work on $switch$ statement of object-oriented language Java.**

Keywords – **action semantics, category theory, equivalence, semantics, Java**

I. Introduction

Theory of categories, which was originally mathematical discipline began to be applied in computer science during the last decades. It is used in various scopes such as algebraic specifications, type theory, automata theory, semantics of programming languages, etc. Categories can express various structures for describing programs. In semantics there are several methods for this task. Each of them has its advantages and its limitations. Action semantics should be solution of their limitations.

So we will focus on these two methods. In the first section we define the basic notions, that are associated with these methods in order to work with them later. In second we prove their equivalence and in the last we provide equivalent expressions of $switch$ statement.

II. Basic Notions

First, we introduce the basic notions in order to work with them later.

A. Category Theory

Category $\mathcal{C}$ is mathematical structure consisting of objects and morphisms. Objects are structures like: sets, groupoids, etc. Morphisms define relations between objects in form $f : A \to B$, where A is domain and B is codomain of morphism f. Morphisms maintain structure of objects and each category must fulfill these conditions [1]:

- every object must have identity morphism: $id_A : A \to A$

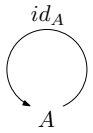


Fig. 1 Identity morphism

- for every two morphisms $f : A \to B$ and $g : B \to C$, where codomain of first morphism is domain of second, must exist their composition: $f \circ g : A \to C$
- for every three morphisms $f : A \to B$, $g : B \to C$ and $h : C \to D$, their composition must be associative: $f \circ (g \circ h) = (f \circ g) \circ h$

Product of objects A, B is object $A \times B$ with projections $\pi_1 : A \times B \to A$ and $\pi_2 : A \times B \to B$. Product is limit if there is exactly one morphism $h : C \to A \times B$ for any object C with morphisms $f : C \to A$, $g : C \to B$, where holds:

$$h \circ \pi_1 = f, \quad h \circ \pi_2 = g$$

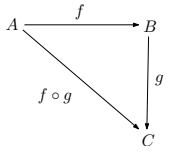


Fig. 2 Composition

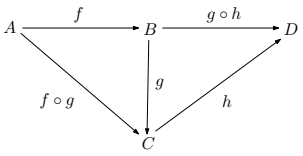


Fig. 3 Associative composition

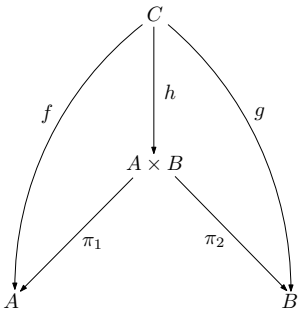


Fig. 4 Product as a limit

B. Action semantics

Action semantics is a semantic method that uses three semantic entities:

- actions
- data
- yielders

Actions are the main type of semantic entities. They present dynamic, computational units to simulate the behavior of the program. A performance of an action may: complete, escape, fail or diverge. Yielders within action present access to data. Data are easily informations processed by actions and they are results of evaluation of yielders. Data and yielders are additional semantic entities. Action notation consists of primitive actions and action combinators. This notation uses English words, so it provides a formal description that is understandable even for readers who have no knowledge about the framework [2]. For example:

execute S and then unfold

where *execute S* and *unfold* are primitive actions and *and then* is an action combinator.

III. Relation of action and categorical semantics

In this section we focus on proving equivalence of action and categorical semantics. We prove it on statements of a simple procedural language in the same way as it is specified in [3][4].

A. Defining of the category

First we need to define the appropriate category:

- ist objects will be states of memory: s_0, s_1, s_2, ...
- its morphisms will be statements of a simple procedural language
- every state s must have identity morphism: $id_s : s \to s$. $Skip$ statement doesn't change the state so it fulfills this condition. We have: $skip : s \to s$
- for every two morphisms $f : s_0 \to s_2$ and $g : s_2 \to s_1$, must exist their composition: $f \circ g : s_0 \to s_1$. This condition is fulfilled by statement sequence $S_1; S_2$, where execution of S_1 is expressed by f and execution of S_2 by g.
- composition is associative, so if we have morphisms $f : s_0 \to s_2$, $g : s_2 \to s_3$ and $h : s_3 \to s_1$ it holds:

$$f \circ (g \circ h) = (f \circ g) \circ h$$
$$f : s_0 \to s_2 \circ (g \circ h) : s_2 \to s_1 = (f \circ g) : s_0 \to s_3 \circ h : s_3 \to s_1$$
$$f \circ (g \circ h) : s_0 \to s_1 = (f \circ g) \circ h : s_0 \to s_1$$

B. Proof of the equivalence

For each statement we must prove two conditions:

1) if $\quad$ **execute**$[\![S]\!]_{s_0} = s_1 \quad \Rightarrow \quad f : s_0 \to s_1$
2) if $\quad f : s_0 \to s_1 \quad \Rightarrow \quad$ **execute**$[\![S]\!]_{s_0} = s_1$

As prove method we use structural induction, which is defined in [3][4].

a) **assignment statement**
Execution of the assignment statement will change the initial state.

1) in action semantics we express it:
execute$[\![x := e]\!]_{s_0}$ =
evaluate$[\![e]\!]_{s_0}$
then
store the given number in the cell bound to x
= s_1

We get state s_1, which is codomain of morphism, that presents assignment statement:
$$f : s_0 \to s_1$$
2) in categorical semantics we express it:
$$f : s_0 \to s_1$$
State s_1 is also final state after execution of assignment statement according to action semantics:
execute$[\![x := e]\!]_{s_0}$ =
evaluate$[\![e]\!]_{s_0}$
then
store the given number in the cell bound to x
= s_1

As we see, both conditions are fulfilled.

b) **skip statement**
This proof is also trivial, we proceed by analogy to the assignment statement.

c) **statement sequence**

1) in action semantics we express it:
execute$[\![S_1; S_2]\!]_{s_0}$ =
execute$[\![S_1]\!]_{s_0}$
and then
execute$[\![S_2]\!]_{s_2}$
= s_1
We formulate inductive assumption:
$$\text{if } \textbf{execute}[\![S_1]\!]_{s_0} = s_2 \quad \Rightarrow f : s_0 \to s_2$$
$$\text{if } \textbf{execute}[\![S_2]\!]_{s_2} = s_1 \quad \Rightarrow g : s_2 \to s_1$$
These assumptions are morphisms whose composition we get an expression of statement sequence:
$$f \circ g : s_0 \to s_1$$
2) in categorical semantics we express it:
$$f \circ g : s_0 \to s_1$$
where:
$$f : s_0 \to s_2,\ g : s_2 \to s_1$$
We formulate inductive assumption:
$$\text{if } f : s_0 \to s_2 \quad \Rightarrow \textbf{execute}[\![S_1]\!]_{s_0} = s_2$$
$$\text{if } g : s_2 \to s_1 \quad \Rightarrow \textbf{execute}[\![S_2]\!]_{s_2} = s_1$$
Both assumptions present subactions of action created by action combinator **and then**:
execute$[\![S_1; S_2]\!]_{s_0}$ =
execute$[\![S_1]\!]_{s_0}$
and then

execute$[\![S_2]\!]_{s_2}$
= s_1

Both conditions are fulfilled.

d) **if statement**
The if statement represents two-way branching. His execution is possible in two ways according to the logical value of conditions. In categorical semantics is if statement expressed by product of states, which is a limit.

1) in action semantics we express it:
execute$[\![if\ b\ then\ S_1\ else\ S_2]\!]_{s_0}$ =
evaluate$[\![b]\!]_{s_0}$
then
check(the given TruthValue is true)
and then **execute**$[\![S_1]\!]_{s_0}$
or
check(the given TruthValue is false)
and then **execute**$[\![S_2]\!]_{s_0}$
= $\begin{cases} s_1 & if\ [\![b]\!]_{s_0} = true \\ s_2 & if\ [\![b]\!]_{s_0} = false \end{cases}$
We formulate inductive assumption:
$$\text{if } \textbf{execute}[\![S_1]\!]_{s_0} = s_1 \quad \Rightarrow f : s_0 \to s_1$$
$$\text{if } \textbf{execute}[\![S_2]\!]_{s_0} = s_2 \quad \Rightarrow g : s_0 \to s_2$$
Inductive assumptions are compositions $h \circ \pi_1$ and $h \circ \pi_2$, where π_1, π_2 are projections of product $s_1 \times s_2$:
$$\pi_1 : s_1 \times s_2 \to s_1, \quad \pi_2 : s_1 \times s_2 \to s_2$$
which is a limit if holds:
$$h : s_0 \to s_1 \times s_2$$

2) like we said, we express if statement by product, which is a limit so there holds:
$$f = h \circ \pi_1,\ g = h \circ \pi_2$$
where:
$$h : s_0 \to s_1 \times s_2,\ \pi_1 : s_1 \times s_2 \to s_1,\ \pi_2 : s_1 \times s_2 \to s_2$$
We formulate inductive assumption:
$$\text{if } f : s_0 \to s_1 \quad \Rightarrow \textbf{execute}[\![S_1]\!]_{s_0} = s_1$$
$$\text{if } g : s_0 \to s_2 \quad \Rightarrow \textbf{execute}[\![S_2]\!]_{s_0} = s_2$$
Both assumptions are subactions which describe the performance of appropriate branches of if statement. If these assumptions holds, there must hold:
execute$[\![if\ b\ then\ S_1\ else\ S_2]\!]_{s_0}$ =
evaluate$[\![b]\!]_{s_0}$
then
check(the given TruthValue is true)
and then **execute**$[\![S_1]\!]_{s_0}$
or
check(the given TruthValue is false)
and then **execute**$[\![S_2]\!]_{s_0}$
= $\begin{cases} s_1 & if\ [\![b]\!]_{s_0} = true \\ s_2 & if\ [\![b]\!]_{s_0} = false \end{cases}$

Both conditions are fulfilled.

e) **while statement**
$While$ statement presents logical cycle with condition at the beginning and unknown number of iterations. We can express it by if statement:
$$if\ b\ then(S;\ while\ b\ do\ S)\ else\ skip$$

1) in action semantics we express it:
execute$[\![while\ b\ do\ S]\!]_{s_0}$ =
unfolding
evaluate$[\![b]\!]_{s_0}$
then
check(the given TruthValue is true)
and then **execute**$[\![S]\!]_{s_0}$

and then **unfold**
or
check(the given TruthValue is false)
and then complete
$$= \begin{cases} s_2 & ak\,[\![b]\!]_{s_0} = true \wedge [\![b]\!]_{s_1} = true \\ s_1 & ak\,[\![b]\!]_{s_0} = true \wedge [\![b]\!]_{s_1} = false \\ s_0 & ak\,[\![b]\!]_{s_0} = false \end{cases}$$
We formulate inductive assumption:
$$\text{if } \textbf{execute}[\![S]\!]_{s_0} = s_1 \quad \Rightarrow g : s_0 \to s_1$$
$$\text{if } \textbf{execute}[\![while\ b\ do\ S]\!]_{s_1} = s_2 \quad \Rightarrow w : s_1 \to s_2$$
$$\text{if } [\![complete]\!]_{s_0} = s_0 \quad \Rightarrow f : s_0 \to s_0$$
Assumptions are compositions:
$$g = h \circ \pi_1, \quad w = f \circ g, \quad f = h \circ \pi_0$$
where π_0, π_1 are projections of product $s_0 \times s_1$:
$$\pi_0 : s_0 \times s_1 \to s_0, \quad pi_1 : s_0 \times s_1 \to s_1$$
which is a limit if holds:
$$h : s_0 \to s_0 \times s_1$$

2) in categorical semantics we express it, so there holds:
$$w = f \circ g, \quad f = h \circ \pi_0, \quad g = h \circ \pi_1$$
where:
$$h : s_0 \to s_0 \times s_1,\ \pi_0 : s_0 \times s_1 \to s_0,\ \pi_1 : s_0 \times s_1 \to s_1$$
We formulate inductive assumption:
$$\text{if } f : s_0 \to s_0 \quad \Rightarrow \textbf{complete} = s_0$$
$$\text{if } g : s_0 \to s_1 \quad \Rightarrow \textbf{execute}[\![S]\!]_{s_0} = s_1$$
$$\text{if } w : s_1 \to s_2 \quad \Rightarrow \textbf{execute}[\![while\ b\ do\ S]\!]_{s_1} = s_2$$
Since the assumptions are fulfilled, there must hold:
execute$[\![while\ b\ do\ S]\!]_{s_0}$ =
unfolding
evaluate$[\![b]\!]_{s_0}$
then
check(the given TruthValue is true)
and then **execute**$[\![S]\!]_{s_0}$
and then **unfold**
or
check(the given TruthValue is false)
and then complete
$$= \begin{cases} s_2 & ak\,[\![b]\!]_{s_0} = true \wedge [\![b]\!]_{s_1} = true \\ s_1 & ak\,[\![b]\!]_{s_0} = true \wedge [\![b]\!]_{s_1} = false \\ s_0 & ak\,[\![b]\!]_{s_0} = false \end{cases}$$

Both conditions are fulfilled.

In every case the conditions were fulfilled so we proved equivalence of action and categorical semantics.

IV. Equivalent expressions of *switch* statement

In this section we provide equivalent expressions of Java *switch* statement with brief description.

Expression of *switch*:

1) in action semantics [5][6]:

execute$[\![switch\ e\ SB]\!]_{s_0}$ =
`evaluate`$[\![e]\!]s_0$
then
furthermore
`case-selectively execute`$[\![SB]\!]_{s_0}$
or
`default`
`trap an unlabeled-break then complete`
$= s_i,\ i = 1 \ldots n$

where:

case-selectively execute$[\![SB: \ case\ C : BS]\!]_{s_0}$ =
`check(the given value is the value of C)`
then
`execute`$[\![BS]\!]_{s_0}$
or
`check(the given value is not the value of C)`
and then
`case-selectively execute`$[\![SB]\!]_{s_0}$

default = `complete | execute`$[\![BS]\!]_{s_0}$

After evaluating the expression e on state s_0 is either the block of $switch$ statement SB, which consists of individual branches, or default branch selectively performed. Selective execution is expressed by composed action **case-selectively execute** which verify whether the value of the expression e matches the value of C for a given case. If they match, then the block of statements BS of given branch is performed. Otherwise, the next case is verified. If there isn't match in any case then the default branch expressed by action **default** is performed. This branch is not mandatory. If it is defined, then it is expressed as performance of block BS, if not, then it is equivalent to action **complete**. Action combiner **_trap_then_** ensures capture of interruptions caused by a $break$ statement and terminates the $switch$ statement.

2) in categorical semantics:

- evaluation of the expression will give us only its value what will not change the state, therefore it is expressed as an empty statement:
$$skip : s_0 \to s_0$$
- more-way branching is expressed by final product of states, so morphism h will have form:
$$h : s_0 \to s_1 \times \cdots \times s_n$$
- projections of the product are:
$$\pi_i : s_1 \times \cdots \times s_n \to s_i, \text{ where } i = 1 \ldots n$$
- so we get compositions, which represent individual branches:
$$f_i = h \circ \pi_i : s_0 \to s_i, \text{ where } i = 1 \ldots n$$
- if the performance of branche is finished by $break$ statement, then $switch$ statement terminates in state obtained by performing this branche
- if the performance of branche is not finished by $break$ statement, then next branche is performed, and this expresses morphism:
$$k_i : s_i \to s_{i+1}, \text{ where } i = 1 \ldots (n-1)$$

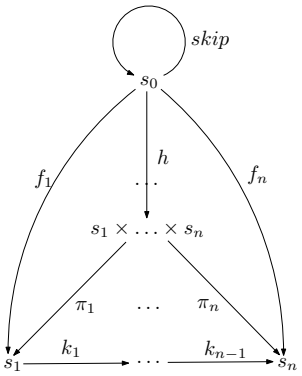


Fig. 5 Categorical expression of $switch$ statement

On the picture 5 we can see that we have two cones there: $\{\pi_i : s_1 \times \cdots \times s_n \to k_i\}$ and $\{f_i : s_0 \to k_i\}$. There is just one morphism $h : s_0 \to s_1 \times \cdots \times s_n$ and there holds these equalities $f_i = h \circ \pi_i$. Therefore, the $switch$ statement represents a categorical limit, whose definition can be found in [1].

V. Conclusion

In this work we introduced the basic principles and properties of action and categorical semantics for description of the programs. In categorical semantics these are objects, which may be different structures, but especially morphisms and their compositions, which define the relationship between objects and describe thereby behavior of the program. This role in action semantics is represented by actions that process data at their execution. These data are provided to them by yielders. Speaking of actions, it may be a primitive action or a compound action formed by the action combinators. They determine the flow of control and informations, in order to gain desired execution of action, which models the behavior of the program. We have proved the equivalence of these methods and we have provided equivalent expressions of Java $switch$ statement. Whereas the action-semantic description of Java is already known, precisely extensions of an equivalent expressions in categorical semantics should be the benefit of this work.

Acknowledgment

This work has been supported by KEGA Grant No. 050TUKE-4/2012:Application of Virtual Reality Technologies in Teaching Formal Methods.

References

[1] V. Novitzká and V. Slodičák, *Kategorické štruktúry a ich aplikácie v informatike*. Košice: Equilibria, 2010.
[2] P. D. Mosses, "Theory and Practice of Action Semantics," *BRICS Report Series*, vol. RS-96-53, December 1996.
[3] H. R. Nielson and F. Nielson, *Semantics with applications*. Wiley and Sons, 1992.
[4] V. Novitzká, *Sémantika programov*. Košice: Elfa, 2001.
[5] D. F. Brown and D. Watt, "JAS: a java* action semantics," *BRICS Report Series*, vol. NS-99-3, March 1999.
[6] ——, "Java action semantics," available on the Internet: http://www.dcs.gla.ac.uk/ daw/publications/JAS.ps.

3D Engines for Visualization of Formally Developed Control Systems

Štefan KOREČKO, Peter NATAFALUŠI

Department of Computers and Informatics, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Slovak Republic

stefan.korecko@tuke.sk, peternataf@gmail.com

Abstract **– In this paper we evaluate three suitable 3D engines that can be used for creation and visualisation of virtual environments. These environments will provide virtual representations of devices, managed by control programs that are developed using formal methods. Virtual representation is needed as the devices are usually not available in physical form, at least not for university teaching. The selected engines are OpenSceneGraph, XNA and Unity 3 and they are evaluated according to their complexity from user's point of view and available scripting languages. Support for virtual reality devices is also taken into account.**

Keywords **– visualization, control program, 3D engine, formal methods**

I. Introduction

With the growing dependency of human society on computer systems, their correctness is becoming an important issue. One of the approaches that can contribute to the correctness is a use of formal methods during design and development of these systems. Formal methods (FM) have strong mathematical background and thanks to this they usually provide some analytical apparatus or can verify certain properties of a system using formal proofs. For example, using methods such as B-Method [1] or Perfect Developer [3], one can write an abstract specification of a software system, including its important properties, and verify that the properties will hold in every state of the system. In addition, he can refine the abstract specification to a concrete, implementable one and proof the correctness of the refinement process. The concrete specification can be automatically translated to a general purpose programming language, such as C, ADA, JAVA or C#.

From the perspective of a teacher one problem of formal methods is that they are not very popular among university students. There are several reasons of this situation. One of them is that FM are harder to learn because of their mathematical nature and, in some cases, because of a need to perform formal proofs manually. And, as we identified in [5], the other important problem is a lack of motivation to learn FM, since students can't sufficiently imagine an appropriate setting where their use is desired and worth the effort: In fact it is not very effective and common to use FM to develop an "ordinary" information system. Their typical domain is a development of safety-critical systems whose failure could result in a loss of lives or valuable assets, such as control software of public transportation or a power plant. It is evident that these typical environments are not available in a university environment. Therefore we proposed [5] to use visualization and simulation software, including game engines, to create virtual representations of the environments and connect formally developed software to them, instead of real devices. In this paper we evaluate three 3D visualisation platforms, or 3D engines - general-purpose OpenSceneGraph and gaming engines XNA and Unity 3. We focus on difficulty of creating virtual environments in them and their support for virtual reality (VR) technologies as one of our goals is to present developed applications using VR equipment available at the Technical University of Košice.

II. OpenSceneGraph

OpenSceneGraph [7] is a graphics application programming interface (API) used for developing applications in areas such as simulation, computer games, virtual reality and modeling. It is written completely in C++ language using OpenGL libraries. Supported operating systems are Microsoft Windows, Mac OS X, Linux, Solaris, FreeBSD to name a few. When working with graphics, the performance of the target machine is crucial. This performance is limited and therefore it is

vital to be very efficient when working with this performance. Since OpenSceneGraph uses C++ and OpenGL and can practically work directly with the hardware, it is able to get a big part of the performance of the machine. OpenSceneGraph is used only do draw a scene, it does not do anything by itself. Being an open source, it can cooperate with various modules that give it additional functionality.

As far as supporting stereo displaying, OpenSceneGraph is able to output various stereo formats. Anaglyphic format is the easiest one, all that is needed are glasses with each glass of a different, contrast color. Blue and red are used most often for this purpose. Quad buffering is another method, however it requires support in the graphics card. It presents different images on one screen, which are separated by using special LCS glasses. These have to be in sync with the display. OpenSceneGraph also supports interleaved image. Now the most beautiful part about all these methods is, that no adjustment of the code is required. The stereo output format is achieved by setting environment variables, or by using command line arguments.

OpenSceneGraph can be integrated with modules called trackers. The purpose of these trackers is to track a 3D input device and read information from its sensors, such as position and rotation. There are many tracker systems, for example OpenTracker, VR Juggler or VRPN. All of these are open source. Each 3D input device that is supported by a given tracker can be used with OpenSceneGraph. It is totally transparent to OpenSceneGraph itself.

The source codes of the API are freely available on the official web site [7]. Thanks to CMake tool, it si possible to compile the source codes on different platforms, although it is a challenging task. There is no dedicated development environment for working with OpenSceneGraph, but it is possible to use Microsoft Visual Studio for instance. Programming language is C++, which uses manual allocation, deallocation and management of memory. While being efficient, it adds to the overall complexity. Taking into account no online documentation, it becomes harder for someone without experience to start using this tool. Moreover, the API makes extensive use of OpenGL. That means, that if we want to start learning OpenSceneGraph, it is required to have a solid knowledge and experience in C++ and OpenGL. It is not possible to learn all three at the same time.

To sum up, OpenSceneGraph is a professional tool for developing graphics applications. As long as a programmer is familiar with it and has some experience, it is an ideal choice to solve any graphical problem. However, to a beginner it would bring more trouble than help.

III. XNA

XNA is a proprietary library for developing games that will run on Microsoft Windows, Xbox 360 and Windows Phone. It is actually an extension of the development tool Microsoft Visual Studio. XNA contains elements specialized on game development, work with graphics and sound. It is based on .Net 2.0 technology. The programming languages are Visual Basic and C#. In theory, every language from the .Net family should be supported, practically only the two mentioned are used in tutorials and templates [8]. In contrast to the previous tool, OpenSceneGraph, XNA is not only able to draw a scene, but also to move object in the scene, to program the logic of the game, its rules and so on. There is no need of any additional module.

However, the API does not include any native support for virtual reality whatsoever, either input nor output. The classes for processing input can work only with devices that support libraries XInput or DirectInput. That means, that if the manufacturer supports these standards, XNA will be able to communicate with them. The classes for input don't include the third dimension, consequently XNA has no option to support 3D input device. The same situation is as for stereo displaying. XNA offers no methods and classes to do this natively. Creating a scene for anaglyphic glasses is possible, though. In short, the scene must be drawn twice and displayed offset. Using a shader draws these scenes and color codes each one for each eye.

Majority of applications with graphical user interface are event driven. Applications developed using XNA are based on a different approach. There is an infinite loop which is executed again and again. Within this loop it is possible to update the game, redraw the scene, process input and so on. Everything in XNA starts with the *Game* class. It offers overridable methods, most important are *Run*, *Update* and *Draw*. Since the programming language is C#, when starting an XNA application, its main method is called. Inside the main method the *Run* method is called, which contains the never-ending loop, in which *Update* and *Draw* are executed.

The official web page for XNA, hosted by Microsoft, offers a complete documentation of XNA API, some tutorials and use cases. There are also many templates and simple games to learn from.

IV. UNITY 3

Unity 3 is another useful tool for creating games. It can be used also to create other applications, like architectural visualization or 3D animation in real time. It comes with its own editor (Fig.

1), which is supported by two platforms so far, Microsoft Windows and Mac OS X. The created application will run on many platforms, including iPhone, Android, iPad. It is even possible to create applications for web browsers, provided these browsers have Unity 3 plug-in installed [9].

As well as XNA, Unity 3 has no native support for stereo displaying nor 3D input devices in its core libraries. However, this can change soon. Virtual Simulations Inc. is considering putting a VRPN plugin for Unity 3 on the market. VRPN is a tracking system for processing input from 3D input devices. It is quite versatile and majority of today's input devices do support it. So once it comes out, Unity will be able to process 3D input very easily, just by integrating the mentioned plug-in.

The applications created by Unity 3 work on a similar bases as XNA. There is an infinite loop which gets executed over and over until the application is closed. But there are some significant differences. Unity has its own editor for creating applications, to begin with. This editor is fully graphical and offers a preview of the scene. That means, we can create the scene, position objects in it, scale them and so on by actually dragging the models in the scene view, scaling them and rotating. It can be done in code too, of course. But the initial scene set up is very straightforward this way. Another priceless feature is the built-in player. It allows to play the game just by pressing one button, no need to build the project beforehand. We can choose an object from the scene to observe, and the player will expose all its variables for reading and writing. This makes debugging very easy.

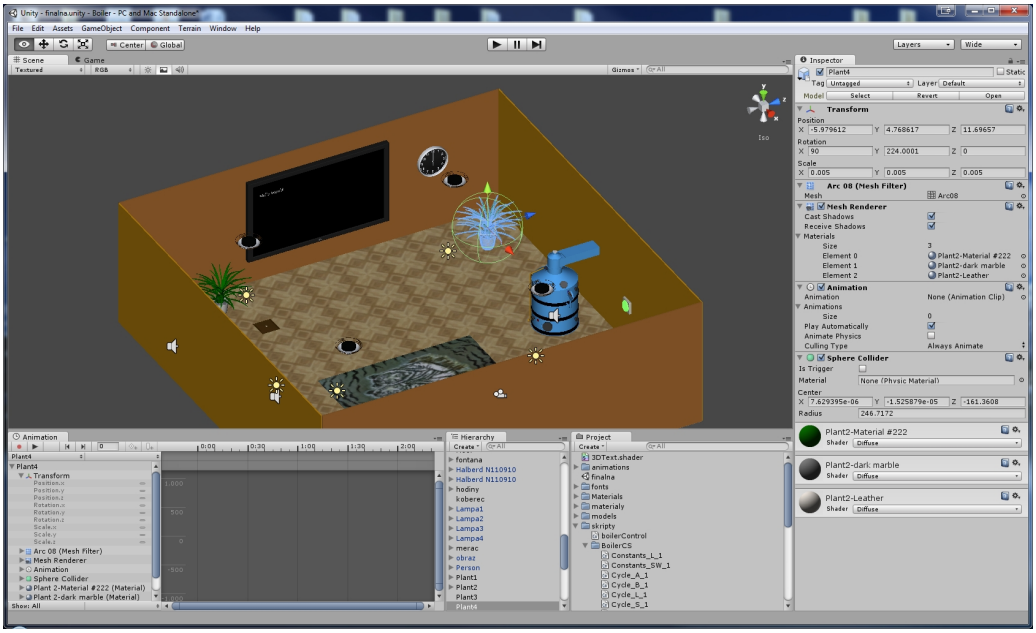


Fig. 1 Unity 3 editor

Unity 3 also uses a different approach to writing the code. In order to program the logic and behavior of the game, it requires to create scripts rather than a single program. These scripts are then associated with objects in the scene. Each script contains overrideable methods and the implementation is up to the programmer. These methods are executed in the appropriate time when the game runs. For instance, the *Update* method is called in each cycle of the infinite loop. Actually, all the *Update* methods of all created scripts are guaranteed to be executed in each cycle. This way, it is very easy and clean to define behavior to objects. These scripts can be written in Javascript, C# and Boo, which is similar to Python.

V. The Most Suitable Engine

Our goal is to build virtual environments, which are to be controlled remotely by external programs, developed using formal methods. The primary criterion for choosing the visualisation tool that should help us to fulfil the goal is an ability to develop the environments as soon and easy as possible. A support for a high-level programming language that is also available as an output language of formal methods tools is also very important. There are two such languages that are supported by Perfect Developer and, via the BKPI [6] compiler, by B-Method - C# and JAVA. XNA as well as Unity 3 do support C# natively and this makes either of them a good choice. Both are also easy to learn and work with, at least when compared with OpenSceneGraph. Unity has a

strong advantage over XNA, which is the dedicated editor. Especially a case, where we want to visualize something rather than to create a functional game. Using the editor, it is very comfortable to build scenes without any knowledge of computer graphics. Of course, when "giving life" to the scene, basic knowledge is a must, but the initial stages can be done very quickly. Next, thanks to the API which supports C#, there is no need to burden the programmer with low level memory management and direct access to the graphics hardware. Everything is encapsulated and the user can start by creating the visualization directly. So, according to our criteria the Unity 3 is a clear choice. It is a pity that OpenSceneGraph is not suitable as it provides the fastest graphics and has the best support for VR equipment. The evaluation summary is provided in Table 1.

Table 1
Summary of important properties of evaluated 3D engines

	OpenSceneGraph	XNA	Unity 3
Programming language	C++	C#	C#
Built-in editor	no	no	yes
Learning curve	steep	easy	easy
Documentation	almost none	MSDN	Scripting reference
User comunity	smaller	large	large
Tutorials	yes	yes	yes

VI. Conclusion

Evaluation of available 3D engines with respect to their potential to support quick and easy development of virtual environment led to the choice of Unity3. The selected engine has been subsequently used to create a scene with a boiler that is managed by a control program developed using B-Method. The scene can be seen in the editor window in Fig. 1. The control program had been adapted from a cased study in [1] and developed in the Atelier-B tool for B-Method. The BKPI compliler [6], developed at the home institution of the authors, has been used to translate the concrete specification of the boiler control program from the language of B-Method to C#. After minor modifications the generated C# code has been integrated with the scene in Unity 3. In future we plan to prepare more virtual environments, new or adapted existing ones, and use them within practices from the subject "Formal specifications of systems".

Acknowledgment

This work has been supported by KEGA grant project No. 050TUKE-4/2012: "Application of Virtual Reality Technologies in Teaching Formal Methods".

References

[1] J. R. Abrial, "The B-Book: Assigning Programs to Meanings," Cambridge University Press, 1996
[2] R.H. Creighton, "Unity 3D Game Development by Example Beginner's Guide," Birmingham, Packt Publishing, 2010. ISBN 978-1-849690-54-6
[3] D. Crocker, "Safe Object-Oriented Software: The Verified Design-by-contract Paradigm" *In: F. Redmill, T. Anderson(Eds.), Practical Elements of Safety, Springer*, 2004, pp. 19-41.
[4] R. Grootjans, "XNA 3.0 Game Programming Recipes: A Problem-Solution Approach," Apress, 2009. ISBN 978-1430218555
[5] Š. Korečko, B. Sobota and Cs. Szabó, "Using Simulation and 3D Graphics Software to Visualize Formally Developed Control Systems," *Proceedings of the 15th WSEAS International Conference on Computers*, Corfu, Greece, 2011, pp. 98-103, ISBN: 978-1-61804-019-0
[6] Š. Korečko, M. Dancák, "Some Aspects of BKPI B Language Compiler Design," Egyptian Computer Science Journal, 35, 3, 2011, pp. 33-43, ISSN 1110-2586
[7] OpenSceneGraph homepage, url: http://www.openscenegraph.org.
[8] XNA homepage, url: http://msdn.microsoft.com/en-us/library/cc178930.aspx.
[9] Unity 3 homepage, url: http://unity3d.com.

The Rôle of Designs in Linear Logic

Pavol MACKO, Valerie NOVITZKÁ, Viliam SLODIČÁK

Department of Computers and Informatics, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Slovak Republic

pavol.macko@tuke.sk, valerie.novitzka@tuke.sk, viliam.slodicak@tuke.sk

Abstract – **Article describes the main notions of ludics, the designs, which are the object corresponding to proofs. A designs captures the goemetrical structure of a sequent calculus derivation. The theory of ludics claims also that proof do not manipulate formulas, but their location, the address where they are stored, what this article also deals. All experiments are described in linear logic, because the theory of ludics and linear logic are closely related.**

Keywords – **linear logic, polarity, ludics, designs**

I. Polarity in linear logic

Polarity of formulas are directly related to the connectives and constants of linear logic, which Girard divided into two classes: *positive* ($\otimes$, $\oplus$, 1, 0, !) and *negative* ($\&$, $\parr$, $\top$, $\bot$, ?) [1], [2], [3]. This distinction has its full meaning in proof-search. Negative connectives are reversible, i.e. the rule to decompose them (bottom-up) is deterministic. Instead, on positive connectives there is a real choice (true non-determinism) [3].

The most important consequence, from the point of view of logic, is that a *cluster* [4] of operations of the same polarity can be seen as a single connective, which is called a synthetic connective. A formula is then an alternation of positive and negative layers.

A formula is positive (negative) if its outer-most connective is positive (negative). In polarized logic, a positive connective can only be applied on positive formulas.

Therefore if P, Q, R are positive formulas, we cannot directly form $((P_1 \parr P_2) \oplus Q^\perp) \otimes R^\perp$. We need use an operation which can change the polarity. In previous work Girard introduced new possibility to change polarity. He described it by using exponentials and he showed that each exponential decomposes into two operations: a core operation which takes care of resources management, and an external shell which changes the polarity. Therefore we have $!N = \downarrow (\sharp N)$ and dually $?P = \uparrow (\flat P)$. This new connection called *Shift* ($\downarrow, \uparrow$) [4], [3]. The role of this new connection Shift is to change the polarity only:

$$\frac{\vdash N, \Gamma}{\vdash \downarrow N, \Gamma} \qquad \frac{\vdash P, \Gamma}{\vdash \uparrow P, \Gamma}$$

Rules mean that if N is negative, then $\downarrow N$ is positive and if P is positive, then $\uparrow P$ is negative.

Now we can write the above formula $(((P_1 \parr P_2) \oplus Q^\perp) \otimes R^\perp)$ using the connective shift in correct form: $(\downarrow (\uparrow P_1 \parr \uparrow P_2) \oplus \downarrow Q^\perp) \otimes \downarrow R^\perp$. In the formula above, we have two synthetic connectives: a negative binary connective $(-\&-)$ and a positive ternary connective $(-\oplus-)\otimes-$.

As we just seen, $A = (P^\perp \oplus Q^\perp) \otimes R^\perp$ can be treated as a single ternary connective $\Phi(P^\perp, Q^\perp, R^\perp)$ which applies to the negative subformulas $P^\perp, Q^\perp, R^\perp$. For this connective there are two possible rules, obtained combining a Tensor-rule with one of the two possible Plus-rules:

$$\frac{\vdash P^\perp, \Gamma \quad \vdash R^\perp, \Delta}{\vdash (P^\perp \oplus Q^\perp) \otimes R^\perp, \Gamma, \Delta}\ A, \{P^\perp, R^\perp\} \quad \text{or} \quad \frac{\vdash Q^\perp, \Gamma \quad \vdash R^\perp, \Delta}{\vdash (P^\perp \oplus Q^\perp) \otimes R^\perp, \Gamma, \Delta}\ A, \{Q^\perp, R^\perp\}$$

Observe that each rule is *labelled* by a pair: the focus and the subformulas which appear in the premises [3].

The dual of A is $A^\perp = (P \& Q) \parr R$, which again is treated as a single ternary connective. The rule for this ternary connective combines the Par-rule with the With-rule [5]:

$$\frac{\vdash P, R, \Lambda \quad \vdash Q, R, \Lambda}{\vdash (P \& Q) \parr R, \Lambda}\ \{(A^\perp, \{P, R\}), (A^\perp, \{Q, R\})\}$$

The rule is labelled by a set of pair: a pair (focus, set of subformulas) for each premise. This set of pair we can write shortly: $(A^\perp, \{\{P, R\}, \{Q, R\}\})$. Negative label is therefore given by the (unique) focus, and a set of subformulas for each premise.

II. Sequent calculus of locations

Next step is to remove logical decorations (formulas). What remains is the location (the address) of the formula. Such an address is called locus. Each formula to be decomposed receives an address, or location, the place where the name of the formula is written. We can assume this space is an infinitely branching tree [6], [5]. If A of the previous section is given the address ξ, then P, Q, R will respectively be distinguished by indices $1, 2, 3$: e.g. Q is the sub-formula of A of relative location 2. P, Q, R will respectively be located in $\xi 1, \xi 2, \xi 3$. The logical rules written above will be interpreted by the decomposition of ξ into the sub-addresses $\xi 1, \xi 2$ and/or $\xi 1, \xi 3$.

The example we have been working with in the previous section can be rewritten as

Positive rules

$$\frac{\xi 1 \vdash \Gamma \quad \xi 2 \vdash \Delta}{\vdash \xi, \Gamma, \Delta}\ \xi, \{1,2\}$$

$$\frac{\xi 2 \vdash \Gamma \quad \xi 3 \vdash \Delta}{\vdash \xi, \Gamma, \Delta}\ \xi, \{2,3\}$$

Negative rules

$$\frac{\vdash \xi 1, \xi 3, \Lambda \quad \vdash \xi 2, \xi 3, \Lambda}{\xi \vdash \Lambda}\ \{(\xi, \{1,3\}, \{2,3\}\})$$

Sequents of addresses are expressions of the form $\Xi \vdash \Lambda$ where: Ξ, Λ are finite sets of addresses, pairwise disjoint; Ξ contains at most one address.

A sequent of the form $\vdash \Lambda$ is said positive, a sequent of the form $\xi \vdash \Lambda$ is said negative. In particular, the empty sequent ($\vdash$) is positive. A sequent is atomic when it contains exactly one address.

Positive rules
Assume I is a set of indices (called ramification [3]), and for $i \in I$ the Γ_i are pairwise disjoint and included in Γ. One can apply the following rule (finite, one premise for each $i \in I$):

$$\frac{\ldots \quad \xi i \vdash \Gamma_i \quad \ldots}{\vdash \xi, \Gamma}\ (\xi, I)$$

Negative rules
Assume N is a set of ramifications (called the directory of the rule), and for all $I \in N$, $\Lambda_I \subseteq \Lambda$. One can apply the following rule (possibly infinite, one premise for each $I \in N$):

$$\frac{\ldots \quad \vdash \xi I \quad \ldots}{\xi \vdash \Gamma}\ (\xi, N)$$

The notation ξI we have used is short for $\{\xi i, i \in I\}$.

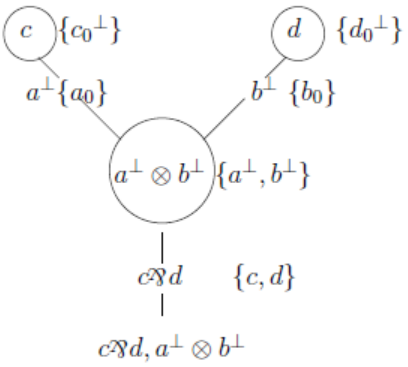
Actions
The pair (ξ, I) is called an *action*[5]. As we have seen, ξ is an address (the address of a formula) and I a set of indices, the relative addresses of the immediate relative subformulas we are considering. ξ is called focus of the action, while I is called ramification.

III. Desings

Designs [5], [4] capture the geometrical structure of sequent calculus derivations. The simplest way to introduce designs is to start from sequent calculus. Let us consider the following derivation, where the rules are labelled by the active formula and the subformulas which appear in the premises:

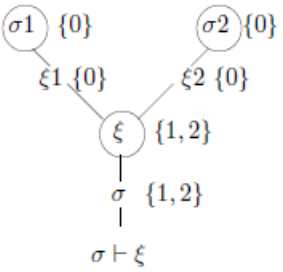
$$\dfrac{\dfrac{\dfrac{\dfrac{\vdash a_0, c_0^\perp}{\vdash a_0, c}\ (c, \{c_0^\perp\})}{\vdash a^\perp, c}\ \{(a^\perp, \{a_0\})\} \quad \dfrac{\dfrac{\vdash b_0, d_0^\perp}{\vdash b_0, d}\ (d, \{d_0^\perp\})}{\vdash b^\perp, d}\ \{(b^\perp, \{b_0\})\}}{\vdash c, d, a^\perp \otimes b^\perp}\ (a^\perp \otimes b^\perp, \{a^\perp, b^\perp\})}{\vdash c \parr d, a^\perp \otimes b^\perp}\ \{(c \parr d, \{c, d\})\}$$

where $a^\perp, b^\perp, c, d$ are formulas which respectively decompose into $a_0, b_0, c_0^\perp, d_0^\perp$. Let us forget everything in the sequent derivation, but the labels. We can see that change of polarity is marked by two types of brackets of labels, negative label $\{...\}$, positive label $(...)$. The derivation above becomes the following tree of labels, which is in fact a (typed) design [5]:



This formalism is more concise than the original sequent proof, but still carries all relevant information to retrieve its sequent calculus counter-part. Positive nodes are marked by circle.

To complete the process, let us now abstract from the type annotation (the formulas), writing only the addresses. In the example above, we locate $a^\perp \otimes b^\perp$ at the address ξ; for its subformulas a and b we choose the sub-addresses $\xi 1$ and $\xi 2$. Finally we locate a_0 in $\xi 10$ and b_0 in $\xi 20$. In the same way, we locate $c\&d$ at the address σ and so on for its subformulas. Our design becomes [5]:



A *design* is given by:

- *a base*, which is a sequent giving the conclusion of the proof (the specification of the process) and
- *a tree of actions* with some properties. A branch in the tree is called a chronicle. If κ_1 is before κ_2 we write $\kappa_1 < \kappa_2$. We think of the tree as oriented in the time, starting from the root. Hence the terminology before and after for actions in the same branch (chronicle).

IV. Slices

The notion of slice was introduced as part of the theory of proofnets in linear logic [3], [7], to speak of the two components of a $\&$-rule. Informally speaking, an $\&$-rule is seen as the super-imposition of two unary rules: $(a\&b, \{a\})$ and $(a\&b, \{b\})$. Given a derivation, we obtain a slice if for any $\&$-rule we select one of the premises. The derivation

$$\dfrac{\dfrac{\dfrac{\vdash \ldots}{\vdash \ldots}\, c}{\vdash a, c}\, a \qquad \dfrac{\dfrac{\vdash \ldots}{\vdash \ldots}\, c}{\vdash b, c}\, b}{\dfrac{\vdash a\&b, c}{\vdash\!\downarrow a\&b, c}\, \downarrow a\&b}\, a\&b, \{\{a\}, \{b\}\}$$

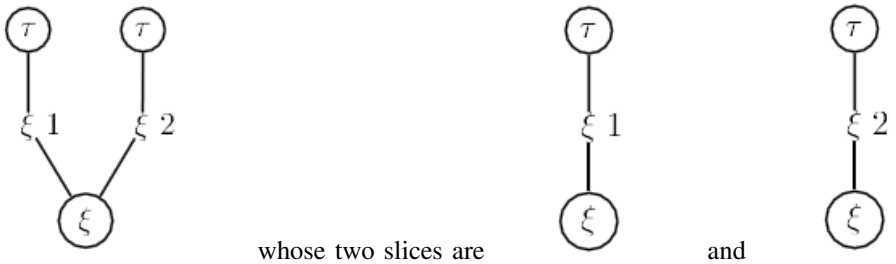
can be decomposed as the super-imposition of two slices (in each slice the $\&$-rule is unary)

$$\dfrac{\dfrac{\dfrac{\vdash \ldots}{\vdash \ldots}\, c}{\vdash a, c}\, a}{\dfrac{\vdash a\&b, c}{\vdash\!\downarrow a\&b, c}\, \downarrow a\&b}\, a\&b, \{a\} \qquad \text{and} \qquad \dfrac{\dfrac{\dfrac{\vdash \ldots}{\vdash \ldots}\, c}{\vdash a, c}\, b}{\dfrac{\vdash a\&b, c}{\vdash\!\downarrow a\&b, c}\, \downarrow a\&b}\, a\&b, \{b\}$$

It is seen that the connective shift has great importance in the construction designs. Observe that selecting one of the premises of a $\&$-rule is exactly what normalization does. Normalization is

always carried out in a single slice: when a negative rule (typically a $\&$-rule) has several premises, the opponent (typically a $\otimes$-rule) selects one of them.

Now we locate c in the address τ, $\downarrow (a\&b)$ in the address ξ, $(a\&b)$ in $\xi 0$, a in $\xi 01$, and b in $\xi 02$. The derivation of previous example corresponds to the following design:

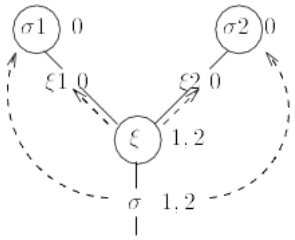


whose two slices are and

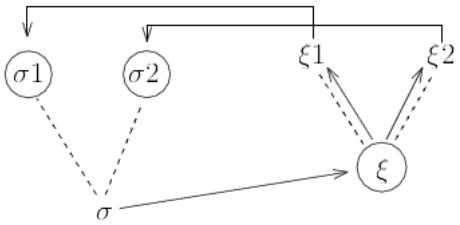
In a design we are given two orders, corresponding to two kinds of information on the actions [5]:

- the succession *in time*, recorded by the chronicles (the chronicles tree);
- the succession *in space*, corresponding to the relation of being sub-address (the prefix tree, which is analogous to a "sub-formula tree").

Now we can again a look at the design in section III. We make explicit the relation of being a sub-address with a dashed arrow connecting σ to $\sigma 1$ and $\sigma 2$, and ξ to its sub-addresses, as follows:



After normalization [5], [3] and treatments our design is follows:



V. Conclusion

In our contribution we presented how to construct designs by using the ludics theory. All sample we are described in linear logic, because the theory of ludics and linear logic are closely related. First problem of of constructing design is polarity, so we closed formulas with same polarity to cluster, which directly describes the change over time. Next step was introduce space into proof by using locations, what is the base of construction design.

The next goal of our research is to develop a theory to describe adding address in the level of proof (quantification of resources at linear logic proof) and find the adequate use of designs in informatics.

References

[1] J. Y. Girard, "Linear Logic," *Theoretical Computer Science*, vol. 50, pp. 1–102, 1987.
[2] J. Y. Girard and Y. Lafont, *Linear logic: Its syntax and semantics*. Cambridge University Press, 2003, pp. 1–42.
[3] J. Y. Girard, *Locus Solum: From the rules of logic to the logic of rules*, 2001, vol. Vol. 11 Issue 3, pp. 301 – 506.
[4] ——, *From foundations to ludics*, 2003, vol. Vol. 9, no. No. 2, pp. 131–168.
[5] C. Faggian and M. Hyland, "Designs, disputes and strategies," S.-V. B. Heidelberg, Ed., 2002, pp. 442–457.
[6] M. Basaldella and C. Faggian, "Ludics with repetitions (exponentials, interactive types and completeness)," 24th Annual IEEE Symposium on Logic In Computer Science, Ed., 2009.
[7] J. Y. Girard, "Proofs and Types," *Cambridge University Press*, pp. 1–175, 2003.

Simple Protection of VoIP Networks Against SPIT Attacks

[1]*Lukáš LAJOŠ,* [2]*Jozef JANITOR*

[1] AT&T Košice, Slovak Republic
[2] Department of Computers and Informatics, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Slovak Republic

[1] lewis39@gmail.com, [2] jozef.janitor@tuke.sk

***Abstract* — The paper is aimed at protecting VoIP networks against SPIT attacks. The aim of this work is a brief analysis and classification of these attacks, but mainly to disassemble different methods and ways how to protect SIP VoIP server. There are also specific methods how to secure VoIP server against the SPIT attacks using Blacklists along with the Audio CAPTCHA application. Application of these methods is carried out in an SIP VoIP server Asterisk. This paper presents specific examples of implementation of the blacklists and adding numbers to the database using the command line as well as through interactive extensions. Finally, there is provided an implementation of Audio CAPTCHA using an external application and technology of Asterisk AGI and is built using PHP programming language.**

***Keywords* — AGI, Asterisk, Audio CAPTCHA, Blacklists, SPIT, VoIP**

I. Introduction to VoIP

This Content of this work is based on the solution of securing SIP VoIP networks against SPIT attacks. VoIP networks are very fast growing technology and security of these networks is on the first place. You can meet this type of communication in every large or international company or even in your own home. In your house it is mainly in the form of communication through some computer programs (e.g. Skype). In business, they are using VoIP phones which look like classic analog phones. The main difference is that they are not analog but digital and they are using the Internet connection for communication in contrast to analog phones which are using POTS (Plain Old Telephone Service) lines for communication. This is the main reason why nowadays this technology so fast growing, because it is cheaper way of communication mainly for international companies which are communicating with branches all over the world. The main pros for this technology are cheap prices for every minute of phone call, quickest installation of new phones than it used to be with old analog lines, more users for one line, etc. The cons include higher costs for initial loading of all devices (VoIP servers, switches, phones, cables) and requirement for technical based staff for installation and setting up of these devices and implementing some security components for VoIP servers, which cause higher costs too.

Despite of these cons there is still the main pro and it is cheaper phone calls in all over the world! As mentioned before, you have the need for implementation of many security components because a lot of attackers currently spread their field of scope to VoIP technology.

A very uncomfortable and disturbing method is SPAM which is known from e-mail communication. In the VoIP it is called SPIT (Spam over Internet Telephony). The goal of this work is to provide some solutions on how to deal with SPIT attacks. Firstly it is very important to understand how this attack works and then we can understand the methods on how to deal with it.

II. SPIT Attacks and How to Deal with Them

A. What is SPIT?

SPIT can be defined as “a bulk unsolicited set of session initiation attempts (e.g., INVITE requests), attempting to establish a voice, video, instant messaging, or other type of communications session”. [1] Although SPIT contains the phrase “SPAM” and has some

parallels with email spam, it also has major differences. The similarity of email spam and SPIT is that in both cases senders (or callers) use the Internet to target recipients (or callees) or a group of users, in order to place bulk unsolicited calls. [2] The main difference between email spam and SPIT is that an email arrives at the email server before it is accessed by the user. This means that structure and content of an email can be analyzed at the server before it arrives at the recipient and so SPAM can be detected before it disturbs the recipient. As in VoIP scenarios delays of call establishment are not wished, session establishment messages are forwarded immediately to the recipients. In other words if the phone rings it is too late for SPIT prevention. Another main difference between email spam and SPIT is the fact that the single email itself contains information that can be used for spam detection. The header fields contain information about sender, subject and content of the message. A single SPIT call in contradiction is technically indistinguishable from a call in general. A SPIT call is initiated and answered with the same set of SIP messages as any other call.

B. How to deal with SPIT?

There are a few methods which can be used to stop SPIT attacks. Five of them are described below:

1) Device Fingerprinting

The technique of active and passive device fingerprinting is based on the assumption that we have knowledge about the type of User Agent that initiates a call. It helps to find out whether a session initiation attempt can be classified as SPIT or not.

2) White Lists, Black Lists, Grey Lists

The White List technique works as follows: each user has a list of users that he accepts calls from and any caller who is not present on the list will be blocked. Black Lists are the contradiction of White Lists and contain only identities that are already known spammers. Any call from a caller whose identity is present in the callee's Black List is blocked. Grey listing works as follows: on initial request of an unknown user (not in White List) the call is rejected and the identity is put on the Grey List. In case the caller tries calling back within a short time period, the call will be accepted. The callee can consult the Grey List and decide, if he will accept future calls from this identity or block it permanently (e.g. put it on the Black List).

3) Reputation Systems

Reputation based mechanisms can be summarized as follows: after receiving a call, the callee can set a reputation value for the caller, that marks this caller as a spitter or not. This reputation value must be assigned to the identity of the caller and can be used for future session establishment requests.

4) Turing tests, Computational Puzzles

Turing tests are tests where the caller is given a challenge that a human can solve easily but that is hard to solve for a machine. Therefore Turing tests or CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart) are tests, that countermeasure Calling Bot attacks in VoIP.

5) Payments at risk

Payments at risk mechanisms can be used in order to demand payment from an unknown caller. In [3] this technique is described as follows: if user A wants to call user B, he must first send a small amount of money to user B. When User B accepts the call and confirms that the call is not a SPIT call, the amount will be charged back to user A. With this technique it is possible to raise costs for SPIT callers while keeping 'normal' calls cheap.

III. USED TECHNOLOGY

A. SIP protocol

SIP is a client-server type of protocol, so communication is the exchange of two types of messages, requests and responses. The client, as well as the server is a logical part of an element. The term 'client' is meant by the phone or software application in use by the user and the term 'server' is referred to as application servers that provide clients with additional services such as registration, location, pricing etc. SIP can be used for other purposes such as network games, instant messaging and more.

B. Asterisk server

From an architectural standpoint, Asterisk is folded from different modules. This modularity usually provides unlimited flexibility to design systems based on Asterisk. An Asterisk administrator can choose which module he wants to use. Each module provides various

functions to the system. For example, a module may allow Asterisk to communicate with analog lines, while others can add the ability to announce calls.

IV. SECURITY SOLUTIONS WITH ASTERISK

A. An interactive list of Blacklist

Through the following code, each user can add a number in the database, remove the number from the database or to add the last number received. It is necessary to add the following context, which will refer to different contexts called by extensions, to the *extensions.conf* file. [8]

```
[from-internal-custom]
exten => *30,1,Goto(custom-blacklist-add,s,1)
exten => *31,1,Goto(custom-blacklist-remove,s,1)
exten => *32,1,Goto(custom-blacklist-last,s,1)
exten => *33,1,Goto(custom-blacklisted,s,1)
```

To add a number to the blacklist there are two ways of doing it. The easiest is to dial *32. You then get a voice telling you the number of the last incoming call. To add it to the blacklist, just press 1. This is great for pesky telemarketers. If you want to add it manually, just press *30. You can then enter the phone number directly on your telephone. To remove a number from the blacklist, press *31 and enter the phone number.

You also need to add a custom function to the *incoming_call* context. This gets called before the normal incoming call handler. So it checks if the incoming number is in the blacklist. This context looks like:

```
[incoming_call]
exten => s,1,LookupBlacklist ; If number blacklisted, goto 102
exten => s,102,Goto(custom-blacklisted,s,1)
```

If the caller number is found in the Blacklist database then this function calls custom-blacklisted context. This context only informs the caller that the callee is not available and the call is terminated.

B. Audio CAPTCHA

Audio CAPTCHA was solved using the AGI script and PHP programming language.

1) Application of the AGI script

In order to execute AGI script, we are required to define an extension in our extensions.conf configuration file, and indicate the execution of our script from there.

```
[from-internal-custom]
include => CAPTCHA
*
[CAPTCHA]
exten => _X.,1,Answer
exten => _X.,2,AGI(CAPTCHA.php)
exten => _X.,3,Hangup
```

Context *[CAPTCHA]* answers the call and invokes AGI script. Further implementation depends on a particular script. When dialplan application takes back execution, the last step will end the call.

The actual PHP script solves the audio CAPTCHA in a simple way, which generates a four-digit number and waits for input from the caller.

The *mtrand* function generates a random four-digit number between 1111 to 9999 and stores it in a variable *$rnum* for further work. The next step will tell generated number and will wait for input from the caller. Then it compares the generated number with input from the user and if not equal, the call ends. If equal, than the application asks the caller to enter the extension number he wants to call.

```
$cnum = 0;
$rnum = mt_rand(1111,9999);
astWrite("SAY NUMBER $rnum #");
astRead();
if ($astOutput != $rnum) astWrite("HANGUP <17>");
if ($astOutput = $rnum) astWrite("SAY NUMBER YOU WANT TO CALL $cnum #");
astCall($cnum);
```

V. Conclusion

SPAM over IP Telephony might be in the future a similar issue as SPAM in emails nowadays. The paper summarized some new methods that can be used to protect the IP Telephony networks against SPIT attacks. A simple installation of the Asterisk PBX and adding either blacklist or captcha features can already today protect IP telephony networks.

Acknowledgment

This work is the result of the project implementation: Development of the Center of Information and Communication Technologies for Knowledge Systems (ITMS project code: 26220120030) supported by the Research & Development Operational Program funded by the ERDF.

References

[1] Dr. Andreas U. Schmidt, Nicolai Kuntze, Rachid El Khayari, SPAM over Internet Telephony and how to deal with it
[2] S. Dritsas, J. Mallios, M. Theoharidou, G.F. Marias and D. Gritzalis, Threat Analysis of the Session Initiation Protocol Regarding Spam, IEEE, 2007.
[3] J. Rosenberg, C. Jennings, RFC 5039 - The Session Initiation Protocol(SIP) and Spam, IETF, 2008.
[4] Asterisk Architecture, The Big Picture, https://wiki.asterisk.org/wiki/display/AST/Asterisk+Architecture+The+Big+Picture, [Available at the Internet: 27.4.2012]
[5] Types of Asterisk Modules https://wiki.asterisk.org/wiki/display/AST/Types+of+Asterisk+Modules, [Available at the Internet: 27.4.2012]
[6] Asterisk: The Defnitive Guide (3rd Edition for Asterisk 1.8), by Leif Madsen, Jim Van Meggelen, and Russell Bryant, Using the Asterisk Database (AstDB), Chapter 10. Deeper into the Dialplan, [Available at the Internet: 27.4.2012] http://www.asteriskdocs.org/en/3rd Edition/asterisk-book-html-chunk/asterisk-CHP-6-SECT-6.html
[7] Nir Simionovich, January 2009, Linux Servers Networking & Telephony Open Source, A Primer to AGI: Asterisk Gateway Interface,http://www.packtpub.com/article/primer-to-agi-asterisk-gateway-interface, [Available at the Internet: 27.4.2012]
[8] 20 thoughts on "Adding blacklist to an Asterisk@Home PBX Server", Adding blacklist to an Asterisk@Home PBX Server, September 26, 2005http://www.jackenhack.com/adding-blacklist-to-an-asteriskhome-pbx-voip-server/, [Available at the Internet: 27.4.2012]
[9] Nir Simionovich, January 2009, Linux Servers Networking & Telephony Open Source, Asterisk Gateway Interface Scripting with PHP, http://www.packtpub.com/article/asterisk-gateway-interface-scripting-with-php,[Available at the Internet: 27.4.2012]

The Role of Digraphs in Modelling Processes

Martina ŽAĽOVÁ , Valerie NOVITZKÁ

Department of Computers and Informatics, Faculty of Electrical Engineering and Informatics Technical University of Košice, Slovak Republic

martina.lalova@tuke.sk, valerie.novitzka@tuke.sk

Abstract **– The research of behavior of processes is very actual theme and brings consecutively new interesting results. We will present an approach of modeling processes - mobile processes in the graphical structure, as digraphs. In our paper we present how digraphs can be constructed for mobile processes of different kind. The mobile processes can be independent structures for relocating and motility can be split into digraphs.**

Keywords **– Mobility, digraphs, actions**

I. Introduction

In computer science the mobile processes are characterized as processes which have changing structure [1], [2].

In these processes the component agents of a system can be linked in various manner and communication between neighbours may carry information or action if there exists a linkage between them. Calculus for mobile processes is based on π-calculus and can express the ability to create and exchange channel names dynamically. Channels are used to execute an action [3].

Is necessary to understand individual aspects and nomenclatures. In programming over wide of mobile area networks new issues has arised, e.g. mobility, the existence of virtual locations, etc [4].

In such systems a new computing paradigm based on the *migration of code* or *agents* is recommended used of the migration of controlses agent.

The research of mobile processes is very interesting and brings many new approaches and results. Our aim is to work out explicit formal definition of our action calculus and action graphs approach of modelling mobile processes based on digraphs.

System of mobile process is made of mobile process package, that are together in a way are in the context. It features inner structure of realization the mobile processes, which make provision for his function. Mobile processes are interacted between self and their behaviour is affected their composition. It is relatively complicate structures.

In our paper we discuss the possibilities of digraphs for modelling mobile processes, labile and motile in one graphical representation and how a digraph of a system can be resolved to a graph for localities and changes.

II. Mobility

Action describes the change in the process, if it affects the intercommunication of several processes [5].

We distinguish two types of shares:

1) action that describes the initial change;
2) reaction, the action is invoked by another action.

Mobility is the possibility of to get to the necessary information, data, or communication from various locations, in different ways, different devices independently of the time and location [6].

Basic aspects of mobility:

1) what moves (process, connection)
2) where does it move:
   - process is moving in real space of computing nodes;
   - process is moving in virtual space connected processes;
   - connection moves in virtual space connected processes.

III. DIGRAPHS

A digraph is directed graph, and it is a diagram composed of nodes and arrows, which going from a node to a node. A *digraph* is a structure consisting of:

- a set of *nodes*,
- a set of *oriented edges*,
- a set of *action*.

In graphical representation the nodes are drawn as circles or ovals. A node represents a locality. Nodes can be similar or they can be nested. The nodes can be linked by edges that are drawn as oriented lines between nodes.

To every node is assigned a *control*, e.g. K or L, where:

- K-node is called *atomic* node, i.e. nothing can be nested within it,
- L-node has always an *arity*, a natural number denoting the number of *ports* at which the edges imping this node.

Because the edges are oriented, there is difference between impinging from inside or outside of a node. An atomic node is obviously passive, because it would make little sense for an atomic control to be active.

The actions are denoted by small letters, e.g. $x, y, \ldots$ and they describe the possible behavior to larger controls.

Action calculus is specified by signature for mobile system S

$$S = (\{U\}, \{\textbf{in}, \textbf{out}, \textbf{rec}, \textbf{open}\} \in M),$$

consisting of the set of nodes U and arity rules **in** and **out** and two special actions **rec** and **open**.

$$\frac{a_1 : m_1 \to n_1, ..., a_r : m_r \to n_r}{K(a_1, ..., a_r) : \quad m \to n} \tag{1}$$

Controls K can include several elements, each of which has input and output arity.

Input arity is defined as the whole number m, output arity is an integer n. r is the 'rank' of controls (number of elements). Grafical representation is on Fig.1.

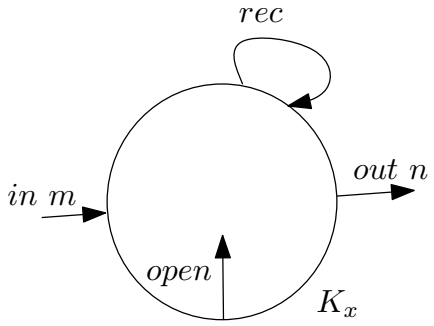


Fig. 1 Action digraphs of controls K with actions

We illustrate simple examples of transition rules in the $action - calculus$ and represent them by digraphs.

We have the rule of control L with agents K_a, K_b, K_y and agent K_n.

```
L: {K_a[ rec P.out a.P | rec P.in a.P |

K_b[ out a.in b.open a |b[out a.in b.open a]]
    | rec Z.open n.( n[out a.in y.open n]
                           | Z)]

| K_y[ rec P.out_ y.P | rec P.in y.P
    | rec Z..open n.( n[out y.in a.open n]
                      | Z)]

| K_n[in a.open n.out a]}
```

where:

- $|$ denotes composition, the agents on the left and the right side act in parallel,
- P is a process,
- $out\ a$ expresses outputting the action a at output port of control K_a,
- $in\ a$ describes inputting of an arbitrary action at the input port of control K_a,
- rec is action of recursion,
- $open$ is action of open control with nesting agent.

To present this rule using digraphs, we need control L with four subcontrols, e.g. agents K_a, K_b, K_y and agent K_n with their arities.

Input arity (2) of control L is $m \to 0$, and at the same time the control terminated - output arity (3) is zero, too. This control serves as a box.

$$\frac{t : m \to 0}{\mathbf{in}(t) \quad m \to 0} \tag{2}$$

$$\frac{s : 0 \to n}{\mathbf{out}(s) \quad n \to 0}, \tag{3}$$

Control K_a used action rec which represented recursion of process P and parallel used action $out\ a$ and $in\ a$ of this process.

Proces P in control K_a used in parallel composition with control K_b.

Control K_b can use action in with parameter b and output action with parameter a and can be $open$ for parameter a.

Behavior of these control is closed on parameter b. After two cycles of execution of the program is behavior already unobservable, because any generate the parameter b, which would be input for the next cycle. This we can see on figure Fig. 2.

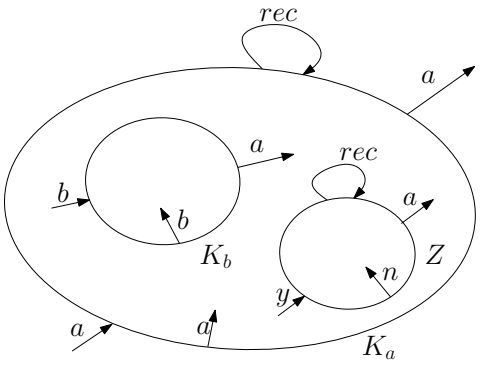


Fig. 2 Digraph of control K_a

To that control K_a appertain part of pseudocode:

```
K_a[ rec P.out a.P | rec P.in a.P |
```

```
K_b[ out a.in b.open a |b[out a.in b.open a]]
    | rec Z.open n.( n[out a.in y.open n]
                              | Z)]
```

Control K_y is recursion process for action rec of process P and there is in parallel action $out\ y$ and $in\ y$ of this process. This process is no the figure Fig. 3.

To that agent appertain part of pseudocode:

```
K_y[ rec P.out_ y.P | rec P.in y.P
    | rec Z.open n.( n[out y.in a.open n]
                         | Z)
```

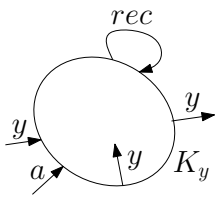


Fig. 3 Digraph of agent K_y

Process P is open for parameter n and control Z.

Control K_a is associated the activities with control K_b and Z.

Processes in control K_n has neutral behavior, it can entering controls and exiting controls, if these controls are open for parameter n. Processes n are nested to the control K_n, illustrated on Fig.4.

To that agent appertain part of pseudocode:

```
K_n[in a.open n.out a]
```

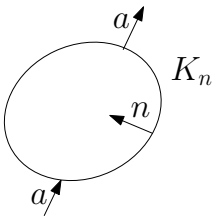


Fig. 4 Digraph of agent K_n

Each controls can move freely in space of control L. The movement of control is not limited. Each control can move in all directions and speed.

If it encounters an obstacle (another process or control) it changes direction. The connection of processes is in move limited by distance of impact and by the input port of the control.

If the connection will be established - the process will associate into the control and it will take over its behavior of control (its direction and speed of movement).

The control describes basic properties of process behavior, which means movement in virtual space. This property defines mobility behavior.

This program system is represented, after minor adjustments, can be visualized using the already existing means of representation to calculus. Similar research has also addressed David Teller and Pascal Zimmer, who designed the application to a graphical representation of π-calculus of mobile processes in the physical model of behavior [7].
Their research consisted of working with ambient system, which is similar to control, but no such predictive value of behavioral processes. They used action entering ambients, exiting ambients and opening ambients, no arities and no complex behavior.

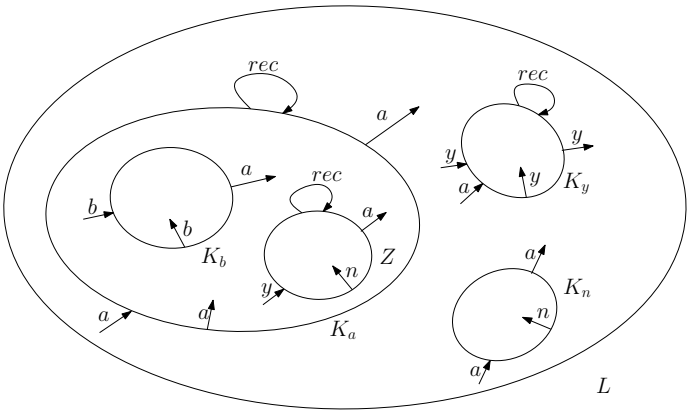


Fig. 5 Digraph of control L with four agents

If we will use the visualization tool for program system designed by us, then we get by subtle adjustments of implementation visualization illustrated on Fig.6.

With this tool we can track the dynamics of mobility. This java application helps in visualization of simple actions and increases the number of processes according to system development, i.e. the elapsed time when the system worked.

This disadvantage becomes clearly visible after a long period of functioning of the system. For more demanding action it is quite unusable, but it could provide good basics for programming of other applications.

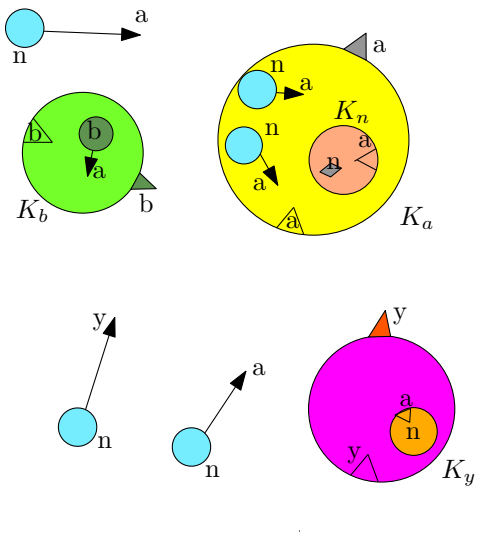


Fig. 6 Example of Control Agents program system L

The new application should include more operations possible as only input and output connections. This we must leave thinking through π-calculi.

Focus on options action calculus, which can be extended and to define the needs action of mobile software systems, that more accurately monitored the properties and behavior of mobile communication.

Mobile communication is defined as a set of applications hardeveru, administrators and users to enable the transmission of data, messages or information according to specified rules, where at least one element of this set is able to relocation - is mobile, has a property of mobility.

As better theoretical-work could was able to serviced also a vision with the principes of category theory, where it has been to possible meditation about establishing own semicategory for single mobile processes and mobile agents.

In [2] are treated special structures, categories as appropriate structures for modelling various complex program systems in rigorous way. There is a *category* consists of a class of *object*s and a class of *morphisms*, that have to satisfy their five basic properties, as is define in the book [2].

Into the class of objects we would interjected processes and into the class of morphism their actions. Syntax of action calculus to us this work considerably will made it easier. As the semantic representation will service action graphs.

IV. Conclusion

In this article we discussed the possibilities of digraphs for representation of process mobility. We used the modified $action\ calculus$ as a syntax and we used action graphs for the graphical representation.

For the description of the system we have chosen significantly abstract structure of four controls which are generating processes and they are merged.

There in our system represented by controls L we have shown several possible scenarios of their behavior. Each control and its process is moving in real space of computing nodes. The process of individual control may terminate its activities, or it can leave the control.

We illustrated simple example of transition rules in the action represented by digraphs. We used the rule of control L with agents K_a, K_b, K_y and agent K_n.
By agent, i.e. control with arity, we define an extended $action\ calculus$, which has axioms to define arity.
Individual axioms of an action in $action\ calculus$ describe the properties of actions which exists in the system and define the behavior of the system.
Action calculus is specified by signature for mobile system consisting of the set of nodes and their arity rules (in and out) and two special actions rec and open. This approach, look on this simplify.
We designed this program system, so that represented the system of four different processes packaged in controls. We used the possibilities of abstraction, so we can better show the behavior of processes and distribution of this behavior through defined actions.
The research of mobile processes is very actual and brings many new approaches and results. Our first aim in the further research is to work out explicit formal definition of our categorical approach of modelling mobile processes based on graphs.

We also would like to investigate the role of mobile processes as the components of the complex program systems. Because we are interesting in the observation of system behaviour, it seem to be useful to use action graphs based on action calculus following the results in [3] for modelling behaviour of mobile processes encapsulated in complex program systems.

Acknowledgment




This work is the result of the project implementation:

Center of Information and Communication Technologies for Knowledge Systems (ITMS project code: 26220120030) supported by the Research & Development Operational Program funded by the ERDF.

References

[1] R. Milner, "Communicating and mobile systems: the pi-calculus," *Cambridge University Press*, 2000.
[2] S. V. Novitzká, V., *Categorical structures and their applications in informatics*. Košice, Slovaika: Equilibria, 2010.
[3] M. P. Slodičák, V., "New approaches in functional programming using algebras and coalgebras," *In European Joint Conferrences on Theory and Practice of Software - ETAPS 2011,Workshop on Generative Technologies, Universität des Saarlandes, Saarbrücken*, p. 1323, March 2011.
[4] G. A. Cardelli, L., "Mobile ambients," *Proc. of FoSSaCS*, vol. Springer LNCS 1378, pp. 140–155, 1998.
[5] L. Cardelli, "Abstraction for mobile computation," *Secure Internet Programming: Security Issues for Mobile and Distributed Objects*, vol. Springer LNCS 1603, pp. 51–94, 1999.
[6] R. Milner, "Elements of interaction," *Communication of the ACM*, vol. 36, pp. 78–89, 1993.
[7] D. Pous, "Les Mobile Ambients en Icobj," internship report (in French). [Online]. Available: http://www-sop.inria.fr/mimosa/ambicobjs/rapport.ps

The Rôle of Designs in Linear Logic

Pavol MACKO, Valerie NOVITZKÁ, Viliam SLODIČÁK

Department of Computers and Informatics, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Slovak Republic

pavol.macko@tuke.sk, valerie.novitzka@tuke.sk, viliam.slodicak@tuke.sk

Abstract – **Article describes the main notions of ludics, the designs, which are the object corresponding to proofs. A designs captures the goemetrical structure of a sequent calculus derivation. The theory of ludics claims also that proof do not manipulate formulas, but their location, the address where they are stored, what this article also deals. All experiments are described in linear logic, because the theory of ludics and linear logic are closely related.**

Keywords – **linear logic, polarity, ludics, designs**

I. Polarity in linear logic

Polarity of formulas are directly related to the connectives and constants of linear logic, which Girard divided into two classes: *positive* ($\otimes$, $\oplus$, 1, 0, !) and *negative* ($\&$, $\parr$, $\top$, $\bot$, ?) [1], [2], [3]. This distinction has its full meaning in proof-search. Negative connectives are reversible, i.e. the rule to decompose them (bottom-up) is deterministic. Instead, on positive connectives there is a real choice (true non-determinism) [3].

The most important consequence, from the point of view of logic, is that a *cluster* [4] of operations of the same polarity can be seen as a single connective, which is called a synthetic connective. A formula is then an alternation of positive and negative layers.

A formula is positive (negative) if its outer-most connective is positive (negative). In polarized logic, a positive connective can only be applied on positive formulas.

Therefore if P, Q, R are positive formulas, we cannot directly form $((P_1 \parr P_2) \oplus Q^\perp) \otimes R^\perp$. We need use an operation which can change the polarity. In previous work Girard introduced new possibility to change polarity. He described it by using exponentials and he showed that each exponential decomposes into two operations: a core operation which takes care of resources management, and an external shell which changes the polarity. Therefore we have $!N = \downarrow (\sharp N)$ and dually $?P = \uparrow (\flat P)$. This new connection called *Shift* ($\downarrow, \uparrow$) [4], [3]. The role of this new connection Shift is to change the polarity only:

$$\frac{\vdash N, \Gamma}{\vdash \downarrow N, \Gamma} \qquad \frac{\vdash P, \Gamma}{\vdash \uparrow P, \Gamma}$$

Rules mean that if N is negative, then $\downarrow N$ is positive and if P is positive, then $\uparrow P$ is negative.

Now we can write the above formula $(((P_1 \parr P_2) \oplus Q^\perp) \otimes R^\perp)$ using the connective shift in correct form: $(\downarrow (\uparrow P_1 \parr \uparrow P_2) \oplus \downarrow Q^\perp) \otimes \downarrow R^\perp$. In the formula above, we have two synthetic connectives: a negative binary connective $(-\&-)$ and a positive ternary connective $(-\oplus-)\otimes-$.

As we just seen, $A = (P^\perp \oplus Q^\perp) \otimes R^\perp$ can be treated as a single ternary connective $\Phi(P^\perp, Q^\perp, R^\perp)$ which applies to the negative subformulas $P^\perp, Q^\perp, R^\perp$. For this connective there are two possible rules, obtained combining a Tensor-rule with one of the two possible Plus-rules:

$$\frac{\vdash P^\perp, \Gamma \quad \vdash R^\perp, \Delta}{\vdash (P^\perp \oplus Q^\perp) \otimes R^\perp, \Gamma, \Delta}\ A, \{P^\perp, R^\perp\} \quad \text{or} \quad \frac{\vdash Q^\perp, \Gamma \quad \vdash R^\perp, \Delta}{\vdash (P^\perp \oplus Q^\perp) \otimes R^\perp, \Gamma, \Delta}\ A, \{Q^\perp, R^\perp\}$$

Observe that each rule is *labelled* by a pair: the focus and the subformulas which appear in the premises [3].

The dual of A is $A^\perp = (P \& Q) \parr R$, which again is treated as a single ternary connective. The rule for this ternary connective combines the Par-rule with the With-rule [5]:

$$\frac{\vdash P, R, \Lambda \quad \vdash Q, R, \Lambda}{\vdash (P \& Q) \parr R, \Lambda}\ \{(A^\perp, \{P, R\}), (A^\perp, \{Q, R\})\}$$

The rule is labelled by a set of pair: a pair (focus, set of subformulas) for each premise. This set of pair we can write shortly: $(A^\perp, \{\{P, R\}, \{Q, R\}\})$. Negative label is therefore given by the (unique) focus, and a set of subformulas for each premise.

II. Sequent calculus of locations

Next step is to remove logical decorations (formulas). What remains is the location (the address) of the formula. Such an address is called locus. Each formula to be decomposed receives an address, or location, the place where the name of the formula is written. We can assume this space is an infinitely branching tree [6], [5]. If A of the previous section is given the address ξ, then P, Q, R will respectively be distinguished by indices $1, 2, 3$: e.g. Q is the sub-formula of A of relative location 2. P, Q, R will respectively be located in $\xi 1, \xi 2, \xi 3$. The logical rules written above will be interpreted by the decomposition of ξ into the sub-addresses $\xi 1, \xi 2$ and/or $\xi 1, \xi 3$.

The example we have been working with in the previous section can be rewritten as

Positive rules

$$\frac{\xi 1 \vdash \Gamma \quad \xi 2 \vdash \Delta}{\vdash \xi, \Gamma, \Delta}\ \xi, \{1,2\}$$

$$\frac{\xi 2 \vdash \Gamma \quad \xi 3 \vdash \Delta}{\vdash \xi, \Gamma, \Delta}\ \xi, \{2,3\}$$

Negative rules

$$\frac{\vdash \xi 1, \xi 3, \Lambda \quad \vdash \xi 2, \xi 3, \Lambda}{\xi \vdash \Lambda}\ \{(\xi, \{1,3\}, \{2,3\}\})$$

Sequents of addresses are expressions of the form $\Xi \vdash \Lambda$ where: Ξ, Λ are finite sets of addresses, pairwise disjoint; Ξ contains at most one address.

A sequent of the form $\vdash \Lambda$ is said positive, a sequent of the form $\xi \vdash \Lambda$ is said negative. In particular, the empty sequent ($\vdash$) is positive. A sequent is atomic when it contains exactly one address.

Positive rules
Assume I is a set of indices (called ramification [3]), and for $i \in I$ the Γ_i are pairwise disjoint and included in Γ. One can apply the following rule (finite, one premise for each $i \in I$):

$$\frac{\ldots \quad \xi i \vdash \Gamma_i \quad \ldots}{\vdash \xi, \Gamma}\ (\xi, I)$$

Negative rules
Assume N is a set of ramifications (called the directory of the rule), and for all $I \in N$, $\Lambda_I \subseteq \Lambda$. One can apply the following rule (possibly infinite, one premise for each $I \in N$):

$$\frac{\ldots \quad \vdash \xi I \quad \ldots}{\xi \vdash \Gamma}\ (\xi, N)$$

The notation ξI we have used is short for $\{\xi i, i \in I\}$.

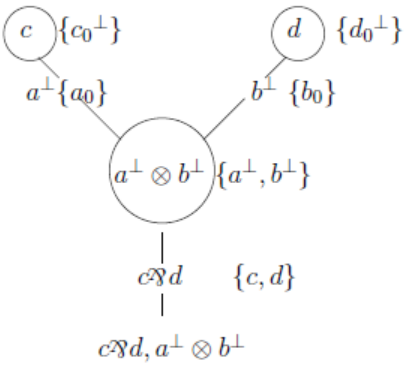
Actions
The pair (ξ, I) is called an *action*[5]. As we have seen, ξ is an address (the address of a formula) and I a set of indices, the relative addresses of the immediate relative subformulas we are considering. ξ is called focus of the action, while I is called ramification.

III. Desings

Designs [5], [4] capture the geometrical structure of sequent calculus derivations. The simplest way to introduce designs is to start from sequent calculus. Let us consider the following derivation, where the rules are labelled by the active formula and the subformulas which appear in the premises:

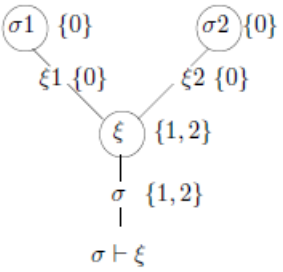
$$\dfrac{\dfrac{\dfrac{\dfrac{\vdash a_0, c_0^\perp}{\vdash a_0, c}\ (c, \{c_0^\perp\})}{\vdash a^\perp, c}\ \{(a^\perp, \{a_0\})\} \quad \dfrac{\dfrac{\vdash b_0, d_0^\perp}{\vdash b_0, d}\ (d, \{d_0^\perp\})}{\vdash b^\perp, d}\ \{(b^\perp, \{b_0\})\}}{\vdash c, d, a^\perp \otimes b^\perp}\ (a^\perp \otimes b^\perp, \{a^\perp, b^\perp\})}{\vdash c \parr d, a^\perp \otimes b^\perp}\ \{(c \parr d, \{c, d\})\}$$

where $a^\perp, b^\perp, c, d$ are formulas which respectively decompose into $a_0, b_0, c_0^\perp, d_0^\perp$. Let us forget everything in the sequent derivation, but the labels. We can see that change of polarity is marked by two types of brackets of labels, negative label $\{...\}$, positive label $(...)$. The derivation above becomes the following tree of labels, which is in fact a (typed) design [5]:



This formalism is more concise than the original sequent proof, but still carries all relevant information to retrieve its sequent calculus counter-part. Positive nodes are marked by circle.

To complete the process, let us now abstract from the type annotation (the formulas), writing only the addresses. In the example above, we locate $a^\perp \otimes b^\perp$ at the address ξ; for its subformulas a and b we choose the sub-addresses $\xi 1$ and $\xi 2$. Finally we locate a_0 in $\xi 10$ and b_0 in $\xi 20$. In the same way, we locate $c\&d$ at the address σ and so on for its subformulas. Our design becomes [5]:



A *design* is given by:

- *a base*, which is a sequent giving the conclusion of the proof (the specification of the process) and
- *a tree of actions* with some properties. A branch in the tree is called a chronicle. If κ_1 is before κ_2 we write $\kappa_1 < \kappa_2$. We think of the tree as oriented in the time, starting from the root. Hence the terminology before and after for actions in the same branch (chronicle).

IV. Slices

The notion of slice was introduced as part of the theory of proofnets in linear logic [3], [7], to speak of the two components of a $\&$-rule. Informally speaking, an $\&$-rule is seen as the super-imposition of two unary rules: $(a\&b, \{a\})$ and $(a\&b, \{b\})$. Given a derivation, we obtain a slice if for any $\&$-rule we select one of the premises. The derivation

$$\dfrac{\dfrac{\dfrac{\vdash \ldots}{\vdash \ldots}\, c}{\vdash a, c}\, a \qquad \dfrac{\dfrac{\vdash \ldots}{\vdash \ldots}\, c}{\vdash b, c}\, b}{\dfrac{\vdash a\&b, c}{\vdash\!\downarrow a\&b, c}\, \downarrow a\&b}\, a\&b, \{\{a\}, \{b\}\}$$

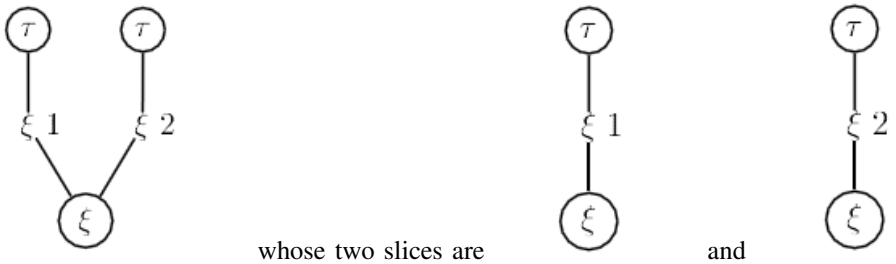
can be decomposed as the super-imposition of two slices (in each slice the $\&$-rule is unary)

$$\dfrac{\dfrac{\dfrac{\vdash \ldots}{\vdash \ldots}\, c}{\vdash a, c}\, a}{\dfrac{\vdash a\&b, c}{\vdash\!\downarrow a\&b, c}\, \downarrow a\&b}\, a\&b, \{a\} \qquad \text{and} \qquad \dfrac{\dfrac{\dfrac{\vdash \ldots}{\vdash \ldots}\, c}{\vdash a, c}\, b}{\dfrac{\vdash a\&b, c}{\vdash\!\downarrow a\&b, c}\, \downarrow a\&b}\, a\&b, \{b\}$$

It is seen that the connective shift has great importance in the construction designs. Observe that selecting one of the premises of a $\&$-rule is exactly what normalization does. Normalization is

always carried out in a single slice: when a negative rule (typically a $\&$-rule) has several premises, the opponent (typically a $\otimes$-rule) selects one of them.

Now we locate c in the address τ, $\downarrow (a\&b)$ in the address ξ, $(a\&b)$ in $\xi 0$, a in $\xi 01$, and b in $\xi 02$. The derivation of previous example corresponds to the following design:

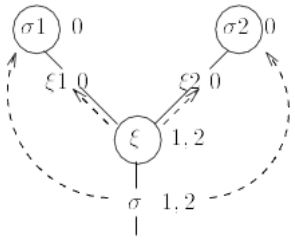


whose two slices are and

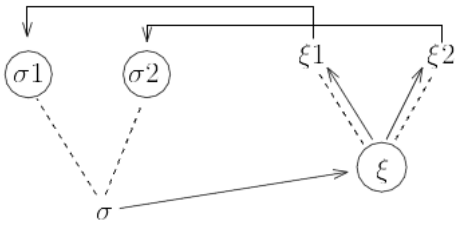
In a design we are given two orders, corresponding to two kinds of information on the actions [5]:

- the succession *in time*, recorded by the chronicles (the chronicles tree);
- the succession *in space*, corresponding to the relation of being sub-address (the prefix tree, which is analogous to a "sub-formula tree").

Now we can again a look at the design in section III. We make explicit the relation of being a sub-address with a dashed arrow connecting σ to $\sigma 1$ and $\sigma 2$, and ξ to its sub-addresses, as follows:



After normalization [5], [3] and treatments our design is follows:



V. Conclusion

In our contribution we presented how to construct designs by using the ludics theory. All sample we are described in linear logic, because the theory of ludics and linear logic are closely related. First problem of of constructing design is polarity, so we closed formulas with same polarity to cluster, which directly describes the change over time. Next step was introduce space into proof by using locations, what is the base of construction design.

The next goal of our research is to develop a theory to describe adding address in the level of proof (quantification of resources at linear logic proof) and find the adequate use of designs in informatics.

References

[1] J. Y. Girard, "Linear Logic," *Theoretical Computer Science*, vol. 50, pp. 1–102, 1987.
[2] J. Y. Girard and Y. Lafont, *Linear logic: Its syntax and semantics*. Cambridge University Press, 2003, pp. 1–42.
[3] J. Y. Girard, *Locus Solum: From the rules of logic to the logic of rules*, 2001, vol. Vol. 11 Issue 3, pp. 301 – 506.
[4] ——, *From foundations to ludics*, 2003, vol. Vol. 9, no. No. 2, pp. 131–168.
[5] C. Faggian and M. Hyland, "Designs, disputes and strategies," S.-V. B. Heidelberg, Ed., 2002, pp. 442–457.
[6] M. Basaldella and C. Faggian, "Ludics with repetitions (exponentials, interactive types and completeness)," 24th Annual IEEE Symposium on Logic In Computer Science, Ed., 2009.
[7] J. Y. Girard, "Proofs and Types," *Cambridge University Press*, pp. 1–175, 2003.

How to Express Denotational Semantics in Categories

Tomáš MARTINKO, Viliam SLODIČÁK

Department of Computers and Informatics, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Slovak Republic

tomas.martinko@student.tuke.sk, viliam.slodicak@tuke.sk

Abstract – **The category theory has been considered a strong tool in mathematics and computer science since its birth. The aim of our work was to use the advantages of the category theory and to apply them in the field of programming language semantics. Individual statements of denotational semantics are displayed in concepts of this theory and there is demonstrated on concrete examples what does this category look like.**

Keywords – **Semantics, programming language semantics, denotational semantics, category theory**

I. Introduction

In 1945 in North America mathematicians Saunders Mac Lane and Samuel Eilenberg established an elegant theory which should have been general enough to explain other fields of mathematics. This theory, which they called the category theory, was developed in the second half of the 20th century and nowadays it is very effectual device in mathematics but mainly in the computer science.
We were motivated to illustrate interconnection between two seemingly isolated branches: the programming language semantics and the category theory where the states are considered to be the objects. We used the method called the denotation semantics for the formal description of the programming language and consequently we expressed the programs inscribed by it in the concepts of the category theory.

II. Denotational semantics

Denotational semantics was created at the beginning of the seventies by Christopher Strachey and Dana Scott. Earlier attempts to formalize the semantics were more focused on the operating meaning of the programs. Unlike those efforts, Ch. Strachey and D. Scott focused on the expression of the semantics of the programming language purely on a mathematical basis. Denotational semantics has brought many benefits. Among others, we can predict the behavior of the program without running it on the computer. Another advantage of denotational semantics is that we can express about features of programs, for example equivalence of two programs.

A. Syntax

Syntactic categories or syntactic domains name collections of syntactic objects that may occur in phrases in the definition of the syntax of the language. Commonly, each syntactic domain has a special metavariable associated with it to stand for elements in the domain. Abstract production rules describe the ways that objects from the syntactic categories may be combined in accordance with the BNF definition of the language. These abstract production rules can be defined using the syntactic categories or using the metavariables for elements of the categories as an abbreviation mechanism[1].

Command ::= while Expression do Command
E ::= N |I |E O E |–E.

B. Semantic

Semantic domains are sets of mathematical objects of a particular form. The sets serving as semantics domains are viewed as normal mathematical sets and structures - for example[1]:

Boolean = {true, false} the set of truth values
Integer = {... , -2, -1, 0, 1, 2, 3, 4, ...} is the set of integers
Store = (Variable → Integer) consists of sets of bindings.

ISBN 978-80-553-0890-6 © 2012 FEI TUKE

C. Connection between syntax and semantics

Semantic functions map objects of the syntactic world into objects in the semantic world. Constructs of the subject language - namely elements of the syntactic domains - are mapped into the semantic domains. These functions are specified by giving their syntax (domain and codomain), called their signatures - for example[1],

$$\text{meaning} : \text{Program} \rightarrow \text{Store}$$
$$\text{evaluate} : \text{Expression} \rightarrow (\text{Store} \rightarrow \text{Value})$$

and by using semantic equations to specify how the functions act on each pattern in the syntactic definition of the language phrases. For example,

$$evaluate[\![E1 + E2]\!]sto = plus(evaluate[\![E1]\!]sto, evaluate[\![E2]\!]sto)$$

states that the value of an expression E1 + E2 is the mathematical sum of the values of its component subexpressions. The value of an expression will depend on the current bindings in the store, here represented by the variable sto. The function evaluate maps syntactic expressions to semantic values - namely, integers - using mathematical operations such as $plus$.

D. Denotation of statements

The result of the execution of the statement in the initial state is a kind of new state, thus after execution the state is changed. Denotation of statement $S \in \mathbf{Statm}$ is defined by semantic function[2]:

$$S_{ds} : Statm \rightarrow (State \rightharpoonup State)$$

which is partialy defined function. Denotations of language statements are defined using denotational equations, while for each alternative of production rule abstract syntax for statements we define one denotation equation. Simple procedural language has five possible forms of statements, which will be defined by the following denotational equations.

Denotation of assignment statement,

$$S_{ds}[\![x := e]\!]s = s[x \mapsto \varepsilon[\![e]\!]s]$$

hence the denotation of assignment statement $x := e$ which is executed in initial state s is new state $s^{'}$. This state differs from the initial state s in value of arithmetic expression e in state s assigned to variable x.

Denotation of empty statement,

$$S_{ds}[\![skip]\!] = id$$

is identical function, what means that the state is not changing.

Denotation of statement sequence,

$$S_{ds}[\![S1; S2]\!] = S_{ds}[\![S_1]\!] \circ S_{ds}[\![S_2]\!]$$

is composed function, defined for the initial state s as follows:

$$S_{ds}[\![S1; S2]\!]s = (S_{ds}[\![S_1]\!] \circ S_{ds}[\![S_2]\!])s$$

Denotation of statement sequence is defined only if it is defined for both statements of sequence.

Denotation of conditional statement,

$$S_{ds}[\![if b then S_1 else S_2]\!] = cond(B[\![b]\!], S_{ds}[\![S1]\!], S_{ds}[\![S2]\!])$$

for initial state s,

$$S_{ds}[\![if b then S_1 else S_2]\!]s = cond(B[\![b]\!], S_{ds}[\![S1]\!], S_{ds}[\![S2]\!])s.$$

Denotation of conditional statement is defined, if it is defined the denotation of particular argument of the auxiliary function $cond$.

Denotation of while statement,

$$S_{ds}[\![while\, b\, do\, S]\!] = cond(B[\![b]\!], S_{ds}[\![S]\!] \circ fix\, F, id).$$

is defined likewise the condition statement by the auxiliary function $cond$, but when $B[\![b]\!] = true$, the denotation of while statement is defined by composition of $S_{ds}[\![S]\!]$ and least fixed point of the functional $fix\, F$.

III. Category theory

Category theory is relatively young branch of mathematics. Yet its influence is being felt in many parts of computer science, including the desing of functional and imperative programing languages, implementation techniques for functional languages, semantics models of programming languages...
It is important to mention, that category theory is not specialized to a particular setting. It is a basic conceptual and notational framework in the same sence as set theory or graph theory, though it deals in more abstract construction and requires somewhat havier notation.

A. Categories

Category $\mathbf{C}$ comprises[3]:

1. a collection of objects;
2. a collection of morphisms;
3. operations assigning to each morphism f an object $dom\, f$, its domain, an object $cod\, f$, its codomain
4. a composition operator assigning to each pair of morphisms f and g, with $cod\, f = dom\, g$, a composite morphism $f \circ g : dom\, f \to cod\, g$, satisfying the associative law;
5. for each object A, an identity morphism $id_A : A \to A$ satisfying the identity law.

B. Products

A product of two objects A and B is an object $A \times B$, together with two projection morphisms $\pi_1 : A \times B \to A$ and $\pi_2 : A \times B \to B$, such that for any object C and pair of morphisms $f : C \to A$ and $g : C \to B$ there is exactly one morphism $< f, g >: C \to A \times B$ making the diagram[4]

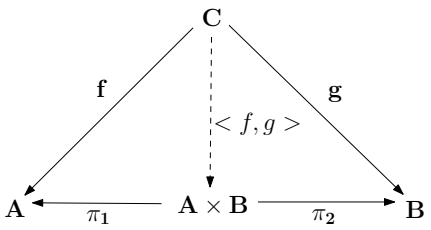


Fig. 1 Product

commute - that is, such that $\pi_1 \circ < f, g >= f$ and $\pi_2 \circ < f, g >= g$.

C. Functors

Let $\mathbf{C}$ and $\mathbf{D}$ be categories. A functor $F : \mathbf{C} \to \mathbf{D}$ is a map taking each $\mathbf{C}$-object A to a $\mathbf{D}$-object $F(A)$ and each $\mathbf{C}$-morphism $f : A \to B$ to a $\mathbf{D}$-morphism $Ff : F(A) \to F(B)$, such that for all $\mathbf{C}$-objects A and composable $\mathbf{C}$-morphisms f and g[4]

1. $F(id_A) = id_{F(A)}$
2. $F(f \circ g) = F(g) \circ F(f)$.

IV. Connection between denotational semantics and category theory

If we comapre denotational semantics and category theory used by categorical semantics, we can see that[5]:

- objects are states;
- morphisms are various language phrases, for example expressions, statements, declarations ...
- set of language phrases is a category, if for each object exists identical morphism, for example for state s it will be the morphism to the same state s;
- for morphisms f and g with $cod\, f = dom\, g$ exists morphism $f \circ g : s \to s_2$ satisfying the associativity law.

A. Basic rules of denotation assignment

In denotational semantics we recognize various basic rules of denotation assignment. Denotations of individual language statements are defined by denotational equations.

Denotation of assignment statement,

$$S_{ds}[\![x := e]\!]s = s[x \mapsto \varepsilon[\![e]\!]s].$$

In category theory it will be morphism from initial state s to the new state $s^{'}$

$$f : s \to s^{'}.$$

Denotation of empty statement,

$$S_{ds}[\![skip]\!] = id$$

will be the morphism from initial state s to the same stae s

$$f : s \to s.$$

Denotation of statement sequence,

$$S_{ds}[\![S1; S2]\!] = S_{ds}[\![S_1]\!] \circ S_{ds}[\![S_2]\!]$$

is a composition of statement S_1 and S_2. These statements must have one commont state and that is also a condition for composition of functions in category theory. Morphism f shows the execution of statement S_1 and morphism g of statement S_2. By composition we get the mapping from state s to state$s^{''}$.

$$f : s \to s^{'},\ g : s^{'} \to s^{''}$$
$$f \circ g : s \to s^{''}$$

Denotation of conditional statement,

$$S_{ds}[\![if b then S_1 else S_2]\!] = cond(B[\![b]\!], S_{ds}[\![S1]\!], S_{ds}[\![S2]\!]).$$

In category theory we will view this statement as product of two objects. Product of objects s_1 and s_2 is object $s_1 \times s_2$ together with two morphisms $\pi_1 : s_1 \times s_2 \to s_1$ and $\pi_2 : s_1 \times s_2 \to s_2$ called projections.

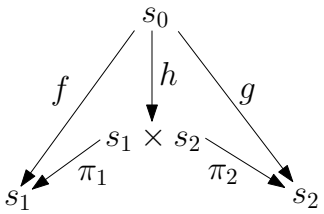


Fig. 2 Product of states

If we compare function $cond(p_1, f_1, f_2)$ and product of objects we can see that the function p is mapping

$$p : s \to s$$

and functions f_1 and f_2

$$f_1 : s \to s^{'}$$
$$f_2 : s \to s^{''}.$$

Denotation of while statement,

$$S_{ds}[\![while\, b\, do\, S]\!] = cond(B[\![b]\!], S_{ds}[\![S]\!] \circ fix\, F, id).$$

will be shown as endofunctor, what is a special case of functor, where the source and target category is the same

$$F : C \to C.$$

B. Examples of categories in denotational semantics

Example 1.

$$\begin{gathered} S[\![z := x; x := y; y := z]\!] = ([x \mapsto 5, y \mapsto 7, z \mapsto 0]) \\ = (S[\![z := x]\!] \circ S[\![x := y]\!] \circ S[\![y := z]\!])([x \mapsto 5, y \mapsto 7, z \mapsto 0]) \\ = S[\![y := z]\!](S[\![x := y]\!](S[\![z := x]\!]([x \mapsto 5, y \mapsto 7, z \mapsto 0]))) \\ = S[\![x := y]\!](S[\![z := x]\!]([x \mapsto 5, y \mapsto 7, z \mapsto 5])) \\ = S[\![z := x]\!]([x \mapsto 7, y \mapsto 7, z \mapsto 5]) \\ = [x \mapsto 7, y \mapsto 5, z \mapsto 5] \end{gathered}$$

Here we have trivial program shown in denotational semantics. It consist from three assignment statements and in brackets at the end we have states. This program will be category if it will contain[5]:

- collection of objects $\Rightarrow$ states s, $s^{'}$, $s^{''}$, $s^{'''}$;
- statements representing the collection of morphisms $\Rightarrow z := x, x := y, y := z$;
- operations that assign to each morphism f an object $dom\, f$ and object $cod\, f \Rightarrow f : s \rightarrow s^{'}$, $g : s^{'} \rightarrow s^{''}$, $h : s^{''} \rightarrow s^{'''}$;
- operator of composition, that assign to each pair of morfphisms f and g, with $cod\, f = dom\, g$, composition morphism $f \circ g : dom\, f \rightarrow cod\, g$ satisfying the associativity law $\Rightarrow f \circ g : s \rightarrow s^{''}$, $g \circ h : s^{'} \rightarrow s^{'''}$;
- for each object A, identical morphism $id_A : A \rightarrow A$ satisfying the identity law $\Rightarrow id_s = S_{ds}[\![skip]\!]_s$

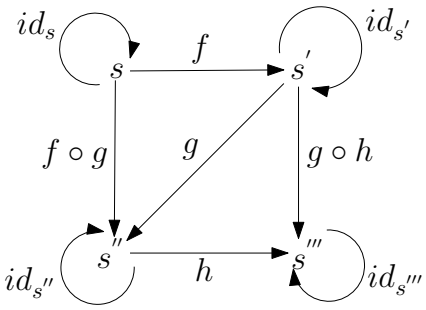


Fig. 3 Example 1. diagram

$$\begin{gathered} f \circ (g \circ h) = (f \circ g) \circ h \\ f : s \rightarrow s^{'} \circ (g \circ h) : s^{'} \rightarrow s^{'''} = (f \circ g) : s \rightarrow s^{''} \circ h : s^{''} \rightarrow s^{'''} \\ f \circ (g \circ h) : s \rightarrow s^{'''} = (f \circ g) \circ h : s \rightarrow s^{'''}. \end{gathered}$$

In this case, the associativity law is satisfied, so $C = (C_{obj}, C_{morp})$ is category.

Example 2.

$$\begin{gathered} S[\![x := 1; if(x = 1)\, then\, y := 2\, else\, skip]\!] = ([x \mapsto 0, y \mapsto 0]) \\ = (S[\![x := 1]\!] \circ S[\![if(x = 1)\, then\, y := 2\, else\, skip]\!])([x \mapsto 0, y \mapsto 0]) \\ = S[\![if(x = 1)\, then\, y := 2\, else\, skip]\!](S[\![x := 1]\!]([x \mapsto 0, y \mapsto 0])) \\ = S[\![if(x = 1)\, then\, y := 2\, else\, skip]\!]([x \mapsto 1, y \mapsto 0]) \\ = cond(B[\![b]\!], S_{ds}[\![y := 2]\!], S_{ds}[\![skip]\!])([x \mapsto 1, y \mapsto 0]) \\ = [x \mapsto 1, y \mapsto 2] \end{gathered}$$

This program consists of one assignment statement and one conditional statement. In terms of category theory this program contains[5]:

- states s, $s^{'}$, $s^{''}$;
- statements representing morphisms $x := 1$, $if(x = 1)\, then\, y := 2\, else\, skip$;
- $f : s \rightarrow s^{'}, g : s^{'} \rightarrow s^{''}, h : s^{'} \rightarrow s^{'}$;
- $f \circ h : s \rightarrow s^{'}$, $f \circ g : s \rightarrow s^{''}$, $i \circ \pi_{s'} = h$, $i \circ \pi_{s''} = g$;
- $id_s = S_{ds}[\![skip]\!]_s$

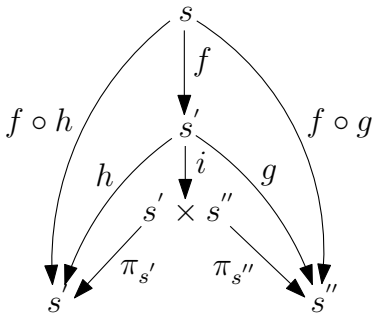


Fig. 4 Example 2. diagram

V. Conclusion

In this work we described considerable deal of category theories with the help of denotation semantics and we demonstrated how exactly the programs can be expressed with the help of category constructions. We went relatively systematically through standard constructions, we adduced their properties and we hope the consequential programs displayed on instances will be interesting for the reader.

Our aim was to isolate key properties of state semantics in the most abstract way. The reason was that in this way we expressed basic structure that we needed for the assigning of the semantics by command. This means that when we moved to more expanded constructions of the simple procedural language we knew what basic elements we had to use for the definition of their semantics.

As mentioned above, the main utilization of the category theory in connection with the programming language semantics is that it affords overview of common properties and relations between denotation domains. This work should serve as the basis for computer experts who would like to analyze relations between more and bigger categories.

Acknowledgment

This work has been supported by KEGA Grant No. 050TUKE-4/2012:Application of Virtual Reality Technologies in Teaching Formal Methods.

References

[1] K. Slonneger, *Formal Syntax and Semantics of Programming Languages*. Reading, Massachuesetts: Addison-Wesley Publishing Company, 1995.
[2] V. Novitzká, *Sémantika programov*. Košice: Elfa s.r.o., 2001.
[3] V. Novitzká and V. Slodičák, *Kategorické štruktúry a ich aplikácie v informatike*. Košice: Equilibria, 2010.
[4] B. Pierce, *Basic Category Theory for Computer Scientists*. Massachuesetts: The MIT Press Cambridge, 2001.
[5] T. Martinko, *Teória kategórií pre sémantiku programovacích jazykov*. Košice: Technická univerzita v Košiciach, Fakulta elektrotechniky a informatiky, 2012.

Time-Spatial Manipulation about Intrusion Detection System Knowledge

Daniel MIHÁLYI

Department of Computers and Informatics, Faculty of Electrical Engineering and Informatics Technical University of Košice, Slovak Republic

Daniel.Mihalyi@tuke.sk

Abstract **– In this contribution we extend our approach related to obtaining knowledge and belief about intrusive attempts from infinite packet streams. A rational agent can obtain rational belief about these unwanted tries from pieces of knowledge based on significant symptoms that are "catched" by intrusion detection system. This system is modelled by the coalgebra of polynomial endofunctor. By using expressive power of resource conscious logic objective knowledge and rational belief from network intrusions can be expressed as a modal formula in resource oriented manner. We call it linear knowledge formula. In our contemporary research we attempt to express how logical time and space can be introduced into the epistemic modal linear formulae by means of manipulation with linear and pitchfork sequent calculus.**

Keywords **– Epistemic logic, Linear logic, Linear sequent calculus, Pitchfork calculus**

I. Introduction

Knowledge and belief can be formulated in Kripke model of possible worlds over coalgebra of appropriate polynomial endofunctor that captures desirable or undesirable behaviour from infinite packet streams. In [1], [2], [3] we modelled intrusion detection system as a coalgebra and we constructed Kripke model for coalgebraic modal linear logic using powerset endofunctor in terms of category theory. In this model we present our idea how a fragment of epistemic modal linear logic can be suitable to provide knowledge and belief about a given intrusion attempt. In this contribution we introduce the idea that relates to extend our existing approach by using time-spatial characteristics for epistemic linear formulae. To do this we formulate in the following steps:

1) first of all we formulate appropriate knowledge linear formula;
2) in the next step we build its proof in linear sequent calculus;
3) afterwards we build a design with logical time that realised by alternation of polarities on it;
4) then we we build time-spatial proof tree in pitchfork calculus;
5) finally we identify loci (addresses) that are (in)comparable by means of relation $\leq$.

II. Linear Sequents and Time-spatial sequents

Linear logic was founded by J.Y. Girard in 1987 [4]. Formulae of linear logic are considered as actions or resources that are produced and consumed in natural way. These resources are treated in logical time/space frame that is determined by relation $\leq$ between loci in locative structure determined by proof tree of pitchfork calculus mentioned in [5], [6]. In this locative structure, when two loci are

- *comparable* w.r.t. relation $\leq$, their relation is timeable;
- *incomparable* w.r.t. relation $\leq$, their relation is spatial, i. e. they are completely independent.

Linear proof tree of a treated formula can be constructed in two steps using

1) linear sequent calculus as a linear deduction system and
2) pitchfork calculus, as a spatial sequent system in sense of [5].

Linear deduction system uses linear sequents written in Gentzen's style either in the right side form or double side form (1)

$$\begin{array}{cc} \vdash \varphi, \psi\Delta & \Gamma, \varphi \vdash \psi, \Delta \\ right\ side\ sequent & double\ sided\ sequent \end{array} \tag{1}$$

where Γ,Δ are linear contexts for formulae φ, ψ.

Time-spatial sequent system uses sequents written in Gentzen's style also as right side or double sequents (2). Every formula in a time-spatial sequent has a locus, i. e. a finite sequence of biases. If we forget everything in a linear sequent (1) except loci of formulae, then we get an expression called pitchfork $\{\xi\} \vdash \Lambda$ where $\{\xi\}$ is a singleton (handle of pitchfork) containing one locus ξ that can be empty set and Λ is a finite set of loci (tines of pitchfork).

$$\begin{array}{cc} \xi \vdash \Lambda & \vdash \Lambda, \xi \\ right\ side\ sequent & double\ sided\ sequent \end{array} \tag{2}$$

where Γ,Δ are spatial contexts for adresser of φ, ψ formulae.

Time-spatial sequent system is constructed in the sense of polarities, where right side sequent rule has positive polarity and double sided rule has negative polarity.

III. Epistemic linear logic for IDS

Epistemic logic is characterised as a logic of objective knowledge and rational belief. Our aim is to show how we can achieve knowledge and belief about intrusion attempts using Kripke model. We use extensional fragment of linear logic and we define the syntax of our epistemic linear logic as follows

$$\varphi ::= \ a_i \mid \varphi_1 \ \& \ \varphi_2 \mid \varphi_1 \oplus \varphi_2 \mid \mathbf{0} \mid \top \mid !\varphi \mid \varphi^{\perp} \mid K_c\varphi \mid B_c\varphi \tag{3}$$

where

- a_i are atomic propositions (i.e. pieces of knowledge),
- $\varphi_1 \ \& \ \varphi_2$ is the extensional conjunction of two formulas φ_1,φ_2,
- $\varphi_1 \oplus \varphi_2$ is the extensional disjunction of two formulas φ_1,φ_2,
- $\mathbf{0}$ is the neutral element of the extensional conjuction.
- $\top$ is the neutral element of the extensional disjunction.
- $!\varphi$ is empiric modal linear operator expressing pleonasm property of formula,
- $\varphi^{\perp}$ is linear negation,
- $K_c\varphi$ denotes that a rational agent c knows that φ,
- $B_c\varphi$ denotes that a rational agent c believes that φ.

We use fragment of epistemic linear logic with epistemic logical operators for objective knowledge and rational belief as a suitable logical system for reasoning about intrusion attempts.

IV. Time-Spatial Preparation for Belief-Desire-Intention Modelling

Using linear logic we show resource oriented manner how beliefs can represent agent informations achieved from Intrusion Detection Epistéme about a "catched" intrusion attempt. Desires also known as linear goals are tasks allocated to the rational agent. Each belief and desire here can be represented as a literal of linear logic. Intentions are implemented as a stack of linear plan instances that are available for a given agent. For this agent the most relevant beliefs we understand as plans that can be realised as a polarized game. These plans are formulated in action sequences of linear logic that are provable in linear sequents and pitchfork sequents consequently. In linear sequents we are able to identify expected polarities and consequently the clusters of linear steps, that can be realised simultaneously in the sense of same polarity based on significant properties of invertibility and focalization. By means of this we can introduce logical time into achieved intentions. In the next step we identify places in linear sequents where polarity has changed and we construct reduced sequents that are assumption for introducing pitchfork calculus, where logical information about formulas are irrelevant and only the locations are important. Pitchfork calculus illustrates how logical space can be well established. In this manner we can illustrate these proofs as polarised Böhm trees that illustrate intentions as a game between player and its adjacent environment.

According to [1] using empiric modality operator we can denote the process from objective knowledge to rational belief as the following linear implication

$$!(K_{007}\varphi \ \otimes \ K_{007}\psi \ \otimes \ K_{007}\tau) \multimap B_{007}\zeta \tag{4}$$

Using linear DeMorgan laws from Table 2 we can rewrite the formula (4) to form

$$!(((K_{007}\varphi)^{\perp} \ \parr \ (K_{007}\psi)^{\perp} \ \parr \ (K_{007}\tau)^{\perp})^{\perp}) \multimap B_{007}\zeta \tag{5}$$

We start with right-side linear sequent of formula (4), i.e.

$$\vdash\ !(((K_{007}\varphi)^{\perp} \parr (K_{007}\psi)^{\perp} \parr (K_{007}\tau)^{\perp})^{\perp}) \multimap B_{007}\zeta, \varphi^{\perp}, \psi^{\perp}, \tau^{\perp}, \zeta^{\perp} \tag{6}$$

and then we build its proof (7) in the sense of linear deduction system by using rules of linear sequent calculus mentioned in Appendix A.

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{}{\tau, \psi, \varphi \vdash \zeta, \varphi^{\perp}, \psi^{\perp}, \tau^{\perp}, \zeta^{\perp}}(id)}{\tau, \psi, K_{007}\varphi \vdash \zeta, \varphi^{\perp}, \psi^{\perp}, \tau^{\perp}, \zeta^{\perp}}(K\!-\!r)}{\tau, K_{007}\psi, K_{007}\varphi \vdash \zeta, \varphi^{\perp}, \psi^{\perp}, \tau^{\perp}, \zeta^{\perp}}(K\!-\!r)}{K_{007}\tau, K_{007}\psi, K_{007}\varphi \vdash \zeta, \varphi^{\perp}, \psi^{\perp}, \tau^{\perp}, \zeta^{\perp}}(B\!-\!r)}{K_{007}\tau, K_{007}\psi, K_{007}\varphi \vdash B_{007}\zeta, \varphi^{\perp}, \psi^{\perp}, \tau^{\perp}, \zeta^{\perp}}((.)^{\perp}\!-\!r)}{K_{007}\psi, K_{007}\varphi \vdash (K_{007}\tau)^{\perp}, B_{007}\zeta, \varphi^{\perp}, \psi^{\perp}, \tau^{\perp}, \zeta^{\perp}}((.)^{\perp}\!-\!r)}{K_{007}\varphi \vdash (K_{007}\psi)^{\perp}, (K_{007}\tau)^{\perp}, B_{007}\zeta, \varphi^{\perp}, \psi^{\perp}, \tau^{\perp}, \zeta^{\perp}}((.)^{\perp}\!-\!r)}{\vdash (K_{007}\varphi)^{\perp}, (K_{007}\psi)^{\perp}, (K_{007}\tau)^{\perp}, B_{007}\zeta, \varphi^{\perp}, \psi^{\perp}, \tau^{\perp}, \zeta^{\perp}}((.)^{\perp}\!-\!r)}{\vdash (K_{007}\varphi)^{\perp} \parr (K_{007}\psi)^{\perp}, (K_{007}\tau)^{\perp}, B_{007}\zeta, \varphi^{\perp}, \psi^{\perp}, \tau^{\perp}, \zeta^{\perp}}(\parr\!-\!r)}{\vdash (K_{007}\varphi)^{\perp} \parr (K_{007}\psi)^{\perp} \parr (K_{007}\tau)^{\perp}, B_{007}\zeta, \varphi^{\perp}, \psi^{\perp}, \tau^{\perp}, \zeta^{\perp}}(\parr\!-\!r)}{((K_{007}\varphi)^{\perp} \parr (K_{007}\psi)^{\perp} \parr (K_{007}\tau)^{\perp})^{\perp} \vdash B_{007}\zeta, \varphi^{\perp}, \psi^{\perp}, \tau^{\perp}, \zeta^{\perp}}((.)^{\perp}\!-\!l)}{!(((K_{007}\varphi)^{\perp} \parr (K_{007}\psi)^{\perp} \parr (K_{007}\tau)^{\perp})^{\perp}) \vdash B_{007}\zeta, \varphi^{\perp}, \psi^{\perp}, \tau^{\perp}, \zeta^{\perp}}(!D\!-\!l)}{\vdash\ !(((K_{007}\varphi)^{\perp} \parr (K_{007}\psi)^{\perp} \parr (K_{007}\tau)^{\perp})^{\perp}) \multimap B_{007}\zeta, \varphi^{\perp}, \psi^{\perp}, \tau^{\perp}, \zeta^{\perp}}(\multimap\!-\!r) \tag{7}$$

We assign polarities to knowledge operators as follows:

- if an epistemic operator occurs on the left side of sequent, its polarity is negative;
- if an epistemic operator occurs on the right side of sequent, its polarity is positive.

that means in the proof tree (7) the formula $B_c\varphi$ is positive and $K_c\varphi$ is negative.

If polarity switching rules are applied in a proof, the polarity is changed. Only the proof steps where polarity is changed are relevant. This leads to the reduced (clustered) form of proof tree (8).

$$\dfrac{\dfrac{\dfrac{\dfrac{\tau, \psi, \varphi \vdash \zeta, \varphi^{\perp}, \psi^{\perp}, \tau^{\perp}, \zeta^{\perp}}{\vdash (K_{007}\varphi)^{\perp}, (K_{007}\psi)^{\perp}, (K_{007}\tau)^{\perp}, B_{007}\zeta, \varphi^{\perp}, \psi^{\perp}, \tau^{\perp}, \zeta^{\perp}}((.)^{\perp}\!-\!r)}{((K_{007}\varphi)^{\perp} \parr (K_{007}\psi)^{\perp} \parr (K_{007}\tau)^{\perp})^{\perp} \vdash B_{007}\zeta, \varphi^{\perp}, \psi^{\perp}, \tau^{\perp}, \zeta^{\perp}}((.)^{\perp}\!-\!l)}{\vdash\ !(((K_{007}\varphi)^{\perp} \parr (K_{007}\psi)^{\perp} \parr (K_{007}\tau)^{\perp})^{\perp}) \multimap B_{007}\zeta, \varphi^{\perp}, \psi^{\perp}, \tau^{\perp}, \zeta^{\perp}}(\multimap\!-\!r) \tag{8}$$

The switch rule $(\multimap\!-\!\mathrm{r})$ we can rewrite by means of Table 2 as follows

$$\dfrac{\dfrac{\dfrac{}{\varphi \vdash \psi, \Delta}(id)}{\vdash \varphi^{\perp}, \psi, \Delta}((.)^{\perp}\!-\!\mathrm{r})}{\vdash \varphi^{\perp} \parr \psi, \Delta}(\parr\!-\!r) \tag{9}$$

At next we have to

- forget all logical relationships between linear formulae including names of linear rules;
- identify linear context Δ as the following sequence $\varphi^{\perp}, \psi^{\perp}, \tau^{\perp}, \zeta^{\perp}$;
- realise substitution of original linear formula $!(((K_{007}\varphi)^{\perp} \parr (K_{007}\psi)^{\perp} \parr (K_{007}\tau)^{\perp})^{\perp}) \multimap B_{007}\zeta$ by α;
- realise substitution of original linear formula; $(K_{007}\varphi)^{\perp}, (K_{007}\psi)^{\perp}, (K_{007}\tau)^{\perp}, B_{007}\zeta$ by β.

After realising these steps we get design, i.e. proof tree consisting of pitchfork form of sequents:

$$\dfrac{\dfrac{\dfrac{\tau, \psi, \varphi \vdash \zeta, \Delta}{\vdash\ \beta, \zeta, \Delta}}{\alpha\ \vdash\ \zeta, \Delta}}{\vdash\ \alpha, \Delta} \tag{10}$$

The proof tree (10) can be regarded as Böhm tree. The steps in this proof express ticktock of logical timer. It means, that we identify proper polarities from original linear proof tree to reduced proof tree as the Böhm tree as a logical time aspect of linear knowledge formula. In logical polarity form (11)

$$\dfrac{\dfrac{\dfrac{-}{+}}{-}}{+} \tag{11}$$

Logical polarity can be illustrated as a game between proponent (polarity $+$) and oponent (polarity $-$) (11).

Table 1
Linear and spatial contexts

Δ	$=$	$\varphi^{\perp}, \psi^{\perp}, \tau^{\perp}, \zeta^{\perp}$
Λ_1	$=$	$\{\xi_1, \xi_2\}$
Λ_{11}	$=$	$\{\xi_{11}, \xi_{12}\}$
Λ_{121}	$=$	$\{\xi_{124}, \xi_{123}, \xi_{122}, \xi_{11}\}$

Linear contexts, where Δ is context of linear calculus and Λ is context of the pitchfork calculus.

Now we introduce space into proof (7). We remove the logical contents Δ of formulae and we introduce spatial contexts according to Table 1. Spatial contexts Δ contain addresses ξ of formulae followed by biases (offsets of immediate subformulae), e.g. $\xi 123$.

Using the biases i.e. fundamental building blocks of locative structure, in (12) we can identify aspects of logical time and space in locative structure of linear knowledge formula.

$$\dfrac{\dfrac{\dfrac{\overline{\xi_{124}, \xi_{123}, \xi_{122} \vdash \xi_{11}, \Lambda_{121}}}{\vdash\ \xi_{12}, \xi_{11}, \Lambda_{11}}\ (+,\vdash \xi_{12},\xi_{11}\{12\})}{\xi_2 \vdash \xi_1, \Lambda_1}\ (-,\xi_1 \vdash \{12,11\})}{\vdash \xi, \Lambda}\ (+,\vdash \xi,\{1\}) \tag{12}$$

For example, biases ξ_{12} and ξ_{124} w.r.t relation $\leq$ are *comparable*, i.e. $\xi_{12} \leq \xi_{124}$ and biases ξ_{123} and ξ_{124} w.r.t this relation are *incomparable*, i.e. $\xi_{12} \leq \xi_{124}$ holds.

V. Conclusion

In our short contribution we present our approach how time-spatial form of objective knowledge and rational belief can be achieved. In the section IV we introduce essential ideas about Belief-Desire-Intention logic model for IDS in resource oriented manner. In the future we would like to extend our approach by formulation of appropriate Kripke frame that related to the Belief-Desire-Intention logic model in time spatial manner.

Appendix A

In this section we introduce linear De Morgan laws and inference rules for linear sequent calculus and spatial pitchfork calculus for presented example.

A. Linear De Morgan laws

Table 2
Linear De Morgan rules [7]

$\mathbf{1}^{\perp} := \perp$	$\perp^{\perp} := \mathbf{1}$
$\top^{\perp} := \mathbf{0}$	$\mathbf{0}^{\perp} := \top$
$(p)^{\perp} := p^{\perp}$	$(p^{\perp})^{\perp} := p$
$\varphi_1 \otimes \varphi_2 := (\varphi_1^{\perp} \parr \varphi_2^{\perp})^{\perp}$	$\varphi_1 \parr \varphi_2 := (\varphi_1^{\perp} \otimes \varphi_2^{\perp})^{\perp}$
$\varphi_1 \,\&\, \varphi_2 := (\varphi_1^{\perp} \oplus \varphi_2^{\perp})^{\perp}$	$\varphi_1 \oplus \varphi_2 := (\varphi_1^{\perp} \,\&\, \varphi_2^{\perp})^{\perp}$
$(!\varphi)^{\perp} := ?\varphi^{\perp}$	$(?\varphi)^{\perp} := !\varphi^{\perp}$
$(\forall x\ \varphi)^{\perp} := \exists x\ \varphi^{\perp}$	$(\exists x\ \varphi)^{\perp} := \forall x\ \varphi^{\perp}$
$\varphi_1 \multimap \varphi_2 := \varphi_1^{\perp} \parr \varphi_2$	

Linear connectives; $\otimes$ = multiplicative conjuction, $\parr$ = multiplicative disjuction, $\&$ = additive conjuction, $\oplus$ = additive disjuction, $\multimap$ = linear implication $(.)^{\perp}$ = linear negation.

B. Linear sequent calculus for proof tree (7)

1) Logical rules::

$$\dfrac{\vdash \varphi, \psi, \Delta}{\vdash \varphi \parr \psi, \Delta}\ (\parr\text{-r}) \tag{}$$

2) exponentials::

$$\dfrac{\Gamma, \varphi \vdash \Delta}{\Gamma, !\varphi \vdash \Delta}\ (!\text{D-l}) \tag{}$$

3) polarity switches::

$$\frac{\varphi \vdash \Delta}{\vdash \varphi^{\perp}, \Delta}\ ((.)^{\perp}-\mathrm{r}) \qquad \frac{\vdash \varphi, \Delta}{\varphi^{\perp} \vdash \Delta}\ ((.)^{\perp}-\mathrm{l}) \qquad \frac{\varphi \vdash \psi, \Delta}{\vdash \varphi \multimap \psi, \Delta}\ (\multimap-\mathrm{r})$$

4) knowledge operators:

$$\frac{\Gamma, \varphi \vdash \Delta}{\Gamma, K_c\varphi \vdash \Delta}\ (\mathrm{K}-\mathrm{l}) \qquad \frac{\Gamma \vdash \varphi, \Delta}{\Gamma \vdash B_c\varphi, \Delta}\ (\mathrm{B}-\mathrm{r})$$

C. Spatial sequents of Pitchfork calculus for proof tree (12)

$$\frac{}{\vdash \Lambda, \xi}\ (\text{✠}) \qquad \frac{\ldots \xi * i \cdots \vdash \Lambda_i \ldots}{\vdash \Lambda, \xi}\ (\xi\vdash\mathrm{N}) \qquad \frac{\cdots \vdash \Lambda_i, \xi * i, \ldots}{\xi \vdash \Lambda}\ (\vdash,\xi,\mathrm{I})$$

ACKNOWLEDGMENT

This work is the result of the project implementation: Center of Information and Communication Technologies for Knowledge Systems (ITMS project code: 26220120030) supported by the Research & Development Operational Program funded by the ERDF.

REFERENCES

[1] D. Mihályi and V. Novitzká, "Towards to the Knowledge in Coalgebraic model IDS," international scientific journal Computing and Informatics, Mar. 2012.
[2] D. Mihályi, V. Novitzká, and M. Ľaľová, "Intrusion Detection System Epistéme," central European Journal of Computer Science.
[3] V. Novitzká and D. Mihályi, "A Coalgebra as an Intrusion Detection System," *Acta Polytechnica Hungarica*, vol. 7, no. 2, pp. 71–79, 2010.
[4] J.-Y. Girard[1], "Linear Logic," *Theoretical Computer Science*, vol. 50, pp. 1–102, 1987.
[5] J.-Y. Girard[3], "From foundations to ludics." pp. 131–168, 2003.
[6] V. Novitzká and D. Mihályi, "Resource-oriented Programming Based on Linear Logic," *Acta Polytechnica Hungarica*, vol. 4, no. 2, pp. 157–166, 2007.
[7] J.-Y. Girard[2], "Linear Logic: Its Syntax and Semantics," in *Advances in Linear Logic (Proc. of the Workshop on Linear Logic, Cornell University, June 1993)*, ser. London Mathematical Society Lecture Note Series, J.-Y. Girard, Y. Lafont, and L. Regnier, Eds., no. 222. Cambridge University Press, 1995.

System for Support of Adaptive Selection of Holiday Destination

Zuzana PANCÁKOVÁ and Viliam SLODIČÁK

Department of Computers and Informatics, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Slovak Republic

zuzana.pancakova@student.tuke.sk,viliam.slodicak@tuke.sk

Abstract – **Selection and selling holiday destinations through the Internet in recent years become a very popular form. An essential means of providing such services to potential customers is a portal that offers the broadest base of holiday offers. In this article we describe a method we have chosen as most appropriate to examine the interests of the customer and optimize the available supply related to holiday destinations. In the event that a particular offer is not exactly coincide with his visions, the system will offer the nearest choice of alternative holiday destination.**

Keywords – **holiday destination, adaptive, click analysis**

I. Introduction

To allow design a system for simplified search holiday destination is necessary to consider the interests of the customer. Currently, there are several ways to determine customer interest. One way is to reach out to customers either directly to the offices of travel agencies or through various means of mass media such as billboards, television and radio advertising, leaflets, internet and other means. Another method is advice from familiar or our own search experience.

Internet is now the preferred form different way to search information. In our case, the search associated with offering holiday destinations. Our portal provides that client offers tailor-made according to his wishes. Use of our website is usable 24 hours a day with the availability of the entire data base offers.

Our criterion is primarily satisfy the customer at any time and as far as possible. If we could not satisfy 100% of his ideas, offer him the closest version approaching its requirements.

II. Goals

In this article we present created web portal for holiday sales, the benefit of which over traditional portal is the ability to adapt to current ideas of the customer. We describe a design method for collecting and analyzing user preferences and use them to see the most preferred holiday destination from the menu.

III. Analysis

A. Holiday destination as object

Holiday destination is somewhat abstract concept, and therefore be described it as an object whose properties acquire any value. In our case, the destination except the name and description defines the following properties: Price, arrival and departure date which values is the definitive information but type of transport, alimentation type , length of stay, location, excursion type, type of accommodation is not possible to quantify.

For these properties it acquired value of the integer constants whose meaning is defined in the code lists. That way it is possible to work with mathematical apparatus.

B. User / customer and their preferences

Any user who is interested in holiday, has a set of preferences on which we ultimately pick the correct and, ideally, it can also purchase. Aim of the system is create him an offer that would best match his imagination. To do so, we must first obtain and store the user preferences.

1) Questionnaire: The first option, as simply and objectively obtain information about user's preferences is to submit the answer sheet. When we defined an object as a holiday destination, the user evaluate percentage of each property corresponding to comply with his personal preferences. This option requires time and attention that is necessary for the user to fill this questionnaire.

2) Click analysis: The method not require user attention, and going to a background. User preference is estimated from user clicks and therefore of interest to a particular destination. Because each specific destination is defined by certain properties, in the case of interest the following characteristics we can to a certain extent transferred to the personal preference of the user. This method, however, for the success work need for start user interaction in the form of a sufficient number of clicks.

C. Choose an appropriate method

Methods of selection and decision-making in general, contains all the rules and procedures. Using the correct method, we can arrive at a choice of holiday destination that offers the greatest compliance with customer's visions, ie adopt the best solution. There is now a wide range of methods of decision making. Currently there are wide range of decision-making methods, but not of all are suitable for solving the problem. To be able to find a separate holiday destinations based on similar features, find apparently hidden connections, we must use some form of multi-criterial decision.

1) The method of decision-making matrix: DMM (Decision Matrix Method) is considered the primary method of multicriteria decision making [7]. It may have more variants. One of the options is based on assessment of weight (importance) of each criteria point scale from 1 to 10 so that level 1 is assigned the smallest weight and gain the greatest level 10. The same scale also assesses the fact that variants selected solutions meet the criteria, ie. level "1" - did not match up to "10" - meets the ideal [3].

The final criterion for the decision shall be elected by the largest weighted sum (the sum of the products of assessment rates to meet the criteria and their weights). Its advantages are simplicity and relatively low process time-consuming, its disadvantages are the high proportion of subjectivity in the valuation of weights in the evaluation criteria and how the various options meet the selected criteria.

2) Our decision algorithm: To implement adaptive destination list was designed a method based on the decision matrix and obtained user preferences to determine compliance with the user's preferences. Based on the classical decision matrix described. This was modified so that it may be used specifically for our case. Decision-making criteria for $(C(D_m))$ in our case consists of all possible destination properties (D_m), such as diet, season, accommodation, transport etc. Each criterion (K_n) states weight, which is a measure of preference (in percentage) for the test user. Zero percent means that the user does not prefer the criterion. One hundred percent on the contrary means full compliance with user preferences. Since not every destination meets all criteria, the value of $x_{m,n}$ shall take the value 1 only if the destination of D_m satisfies the criterion of K_n. About it as a given destination meets the user's preferences deals sum of all $x_{m,n}$ for a given destination S_m, each $x_{m,n}$ is multiplied by the weight of P_n. Destination that receives the highest value of S, obtained the highest score and thus comes closest to the user's preferences. The following table shows the revised decision matrix used in our application 1.

Table 1
Table decision-making matrix used in our application

	p	D_1	D_2	D_3	D_4	$\cdots$	D_m
K_1	P (K_1)	X (K_1, D_1)	X (K_1, D_2)	X (K_1, D_3)	X (K_1, D_4)	$\cdots$	X (K_1, D_m)
K_2	P (K_2)	X (K_2, D_1)	X (K_2, D_2)	X (K_2, D_3)	X (K_2, D_4)	$\cdots$	X (K_2, D_m)
K_3	P (K_3)	X (K_3, D_1)	X (K_3, D_2)	X (K_3, D_3)	X (K_3, D_4)	$\cdots$	X (K_3, D_m)
K_4	P (K_4)	X (K_4, D_1)	X (K_4, D_2)	X (K_4, D_3)	X (K_4, D_4)	$\cdots$	X (K_4, D_m)
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
K_n	P (K_n)	X (K_n, D_1)	X (K_n, D_2)	X (K_n, D_3)	X (K_n, D_4)	$\cdots$	X (K_n, D_m)
$\sum X$	$-$	S_1	S_2	S_3	S_4	$\cdots$	S_m

p - weight; p_m = P (K_n); $p \in < 0, 100 >$
D_m - destination
K_n - criterion; k_n = K (D_m)
x - value; $x_{m,n}$ = X (K_n, D_m); $x \in < 0, 1 >$
$S_m = \sum_{i=1}^{n} X(K_n, D_m)$

IV. Web application

Web application that uses the above-mentioned principles and practices and was developed in PHP5 programming language with object-oriented approach. The entire data base application is stored in MySQL database type.

A. User

In order to apply the mentioned algorithms, the application must distinguish between users. The distinction is ensured by the mandatory registration and user login. This approach is somewhat bothering the user, but one of the most effective method.

B. Questionnaire

The user after login has ability to fill a questionnaire that captures and stores the users' interests and preferences. After questionnaire 1 completing the application can already provide a list of destinations to which the order is modified according to user's preferences.

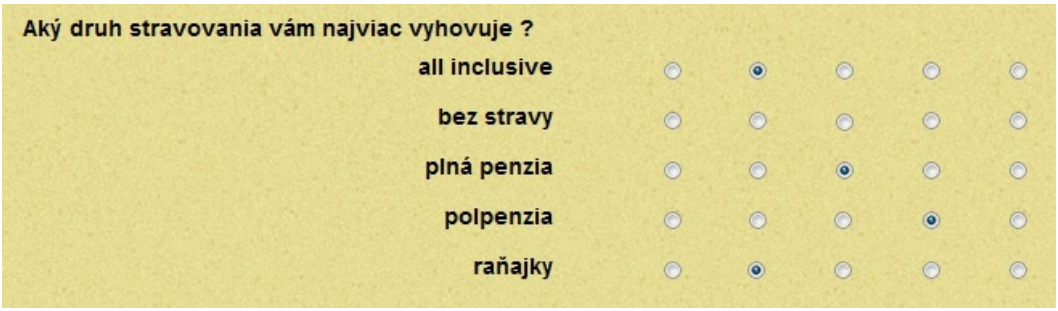


Fig. 1 Web application - questionnaire

The application allows display user preferences 2, and preference take their values $< 0, 100 >$ and we view them as a percentage. The value indicates how much percent corresponds to the preference given to the user's imagination. The default value for each preference is 50%, which represents a compromise between the least preferred and most preferred choice.

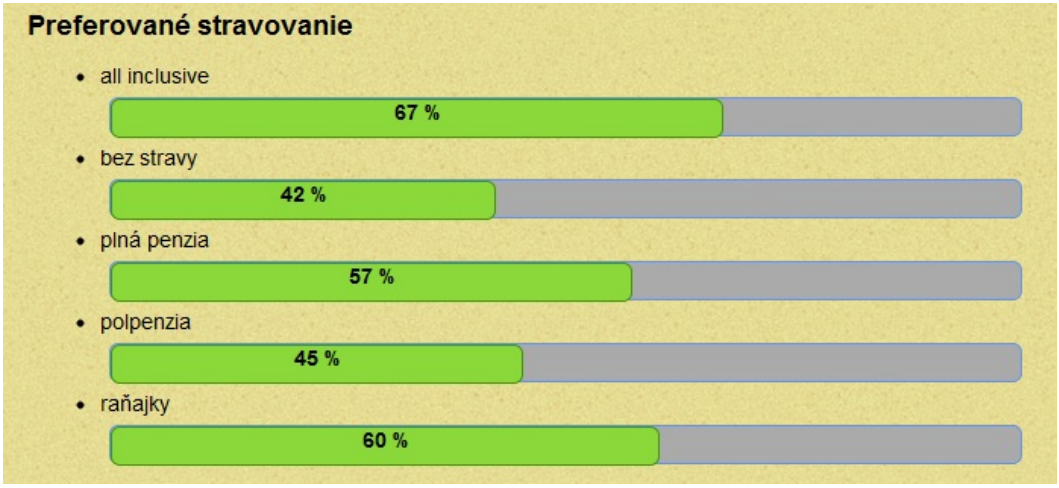


Fig. 2 Web application - user preferences only from questionnaire

C. Holiday destination offer

Databases include holiday destinations that appear in the comprehensive list immediately after login 3. The order of this list is different for each user and depends on the actual preferences of the user. We assume that the travel destinations that appear among the first at the top of the screen are seen first by the user. If we present travel destination of which the user could have an eventual interest, we can rapidly increase the chance of one of these destinations and purchase.

D. User preferences

This module is used for graphical representation of a single analysis for each criterion. Here we have the opportunity to see and compare the extent to which the results of the questionnaire and click analysis differ 4.

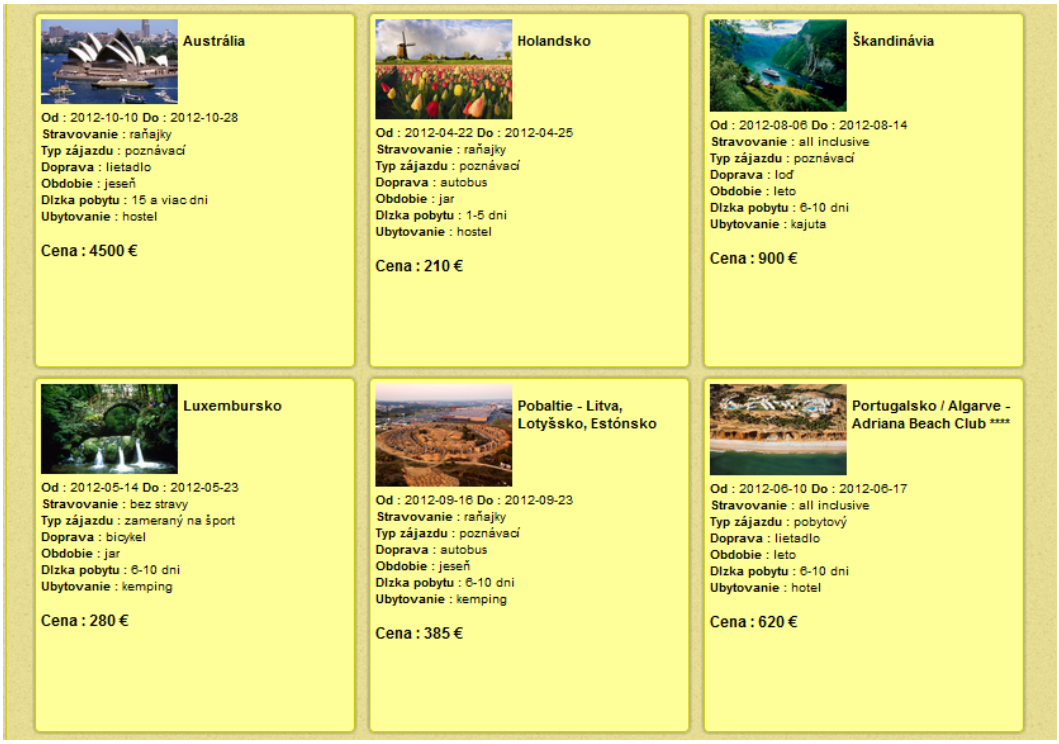


Fig. 3 Web application - offer

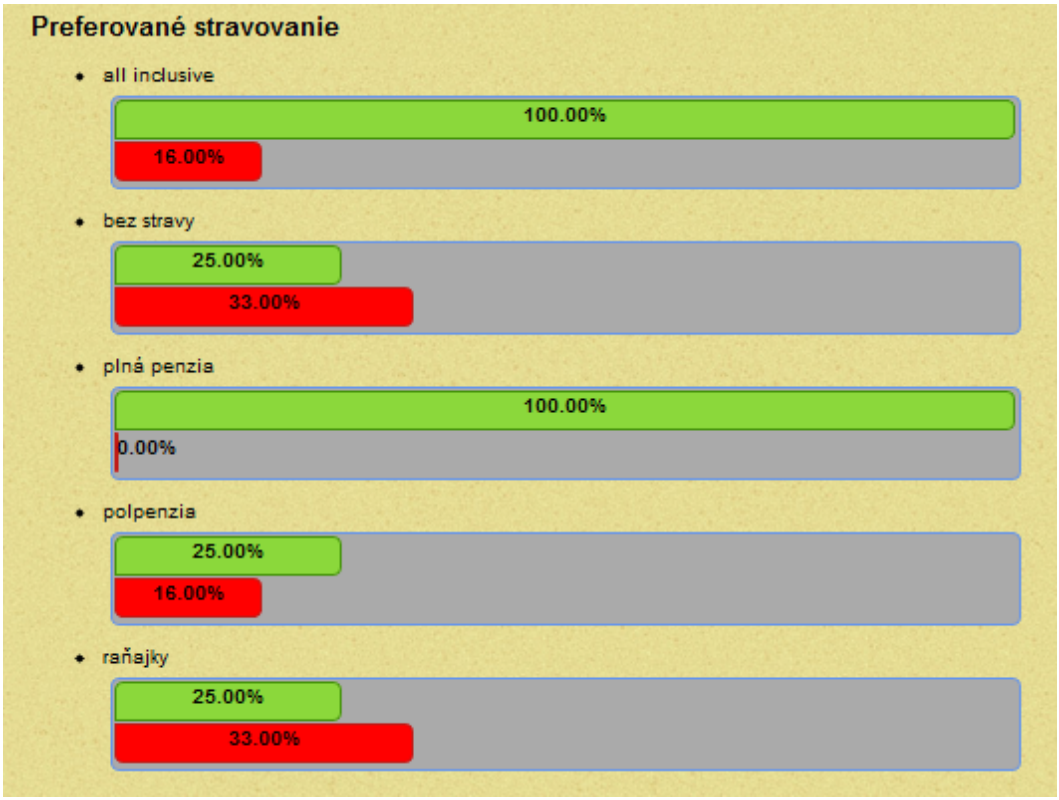


Fig. 4 Web application - user preferences

E. Export to PDF

PDF files can contain text and pictures. In our case the PDF file serves to store all offers, even with the relevant pictures. In the event that the user is logged off in the application, may at any time open a PDF file with all destinations, which have previously been exported to a PDF file and saved somewhere.5.



Fig. 5 Export offers to PDF

F. Customer profile vs. results of the systems

If the system is working properly, the real preferences of the customer preference and analyzed by our system should be similar in certain areas. Even the questionnaire is a great form of getting real customers' preferences. For it was also included in the application, but not every customer has to fill in the questionnaire (because it is optional) and there is no guarantee that the data entered is consistent with the actual idea.

An example of evaluation results our customer :

1) Customer **Verbal formulation preferences :**
   He is interested mainly for sightseeing tours.
   Results : 6

G. Success

The success of the system is that if most of the results of the completed questionnaire are almost identical to the results of click analysis. The more often users change their views when completing the questionnaire and also the more often they click on the destination, the more probable it is that there will be more consistency between the two graphs. The findings are that we can say : "Precision of results of click analysis is directly proportional to the number of clicks on the destination that the user should do." Another truth is, that "accuracy also increases the diversity of choice of holiday destination" 7.

V. Consolution

The goal of this work was to design and create a system that allows a usercustomer offers to him the most interesting holiday destinations, thereby increasing their chance to buy. This feature gives operator of this system a significant marketing advantage. Was created a web portal which uses the previously mentioned principles. Includes administration interface that allows the addition of new holiday destinations and also user management with their preferences. The user interface is available after the registration and provides an offer of possible holiday destinations and questionnaire. Current preferences of the user can view or modify at any time.

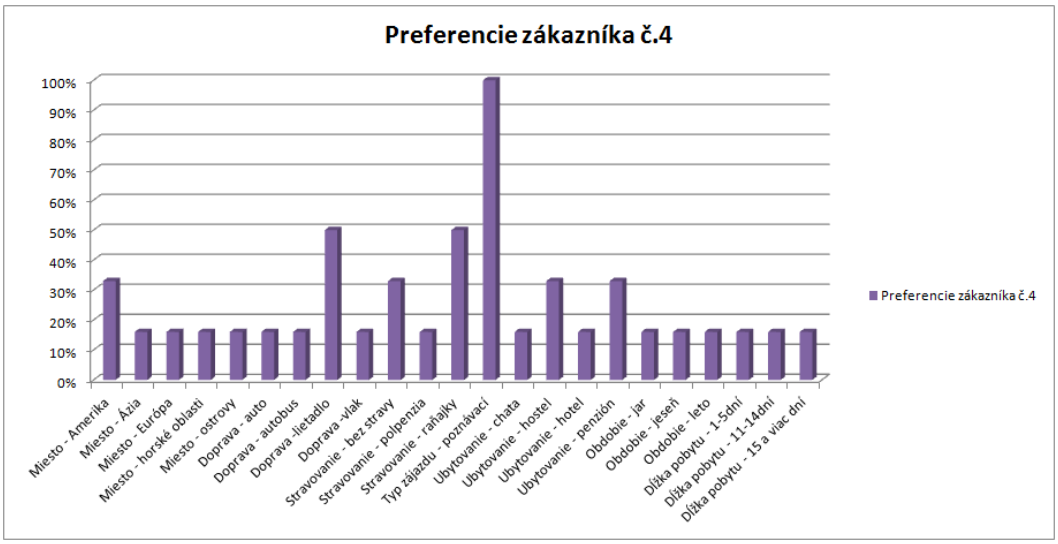


Fig. 6 Resulting graph preferences

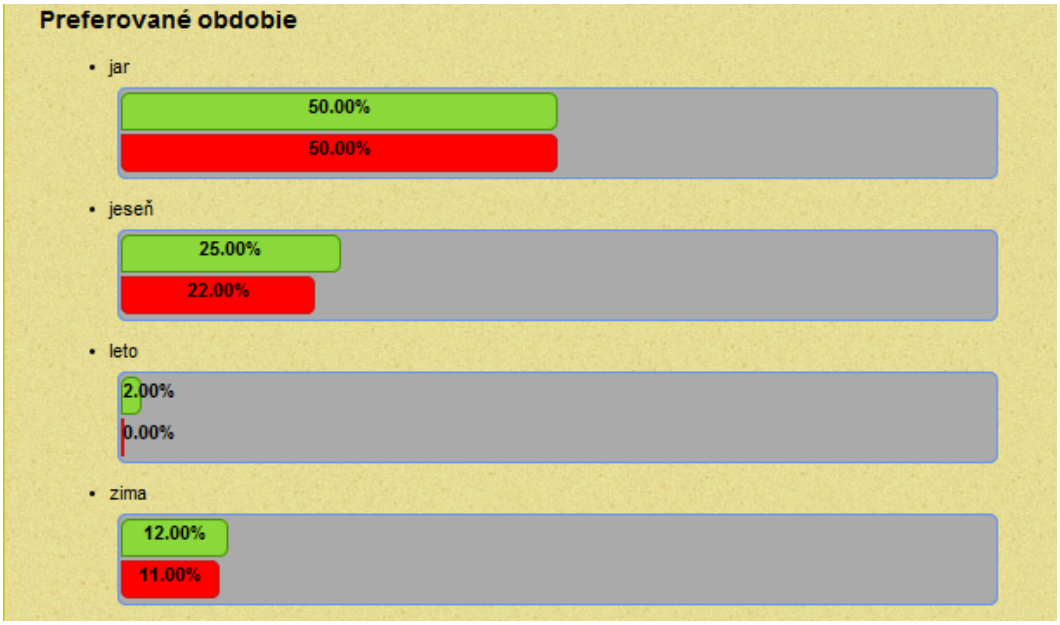


Fig. 7 Success

ACKNOWLEDGMENT

This work has been supported by KEGA Grant No. 050TUKE-4/2012:Application of Virtual Reality Technologies in Teaching Formal Methods.

REFERENCES

[1] Matteucci, M.: *A Tutorial on Clustering Algorithms.*
[2] Magdolenová, J.: *Empirické metódy rozhodovania v manažmente.*
[3] Leitner,B.: *Viackriteriálne multikriteriálne rozhodovanie.Rozhodovacia analýza.*
[4] Ocelíková, E.: *Viackriteriálne rozhodovanie, Košice, 2011.*
[5] Řezanková, H. - Húsek, D. - Snášel, V.: *Shluková analýza dat, Professional Publishing, Praha, 2007.*
[6] Hrouda, J.: *Hodnocení alternativ metodou relativních významností, Ekonomicko-matematický obzor 19, 1983, č.4 .*
[7] Wang, H. - Xie, M. - Goh, T. N.: *Total Quality Management and Business Excellence, 2003 .*

The SLA Meter Tool

Adrián PEKÁR, Juraj GIERTL, Martin RÉVÉS, Peter FECIĽAK and Miroslav ANTL

Dept. of Computers and Informatics, Faculty of Electrical Engineering and Informatics
Technical University of Košice, Slovak Republic

adrian.pekar@cnl.tuke.sk, juraj.giertl@cnl.tuke.sk, martin.reves@cnl.tuke.sk,
peter.fecilak@cnl.tuke.sk, miroslav.antl@cnl.tuke.sk

Abstract **– This paper provides an insight into the design and development of the SLA Meter tool. SLA Meter is being developed by the MONICA research group at the Computer Networks Laboratory at the Technical University of Košice. The tool enables to check whether the specified conditions between the subscriber and the provider are met. The design of the SLA Meter was based on the BasicMeter tool and as its predecessor, it also consists of several parts at different layers.**

Keywords **– Modular Web Application, Network Monitoring, SLA Evaluation, SLA Meter**

I. Introduction

A large part of the earth's population became a user of the Internet and other network services. Networks are becoming a part of the daily life, whether it is at work, during spare time or in new forms such as parts of an intelligent home. This results in the rapid development of the networks and their utilization. Subsequently, not only the number of users, but also the number of providers is growing and the mutual control of the pre-agreed conditions is becoming even more demanding.

Service Level Agreement (SLA) is dedicated to the creation and check-up of agreements about the provided services. This field of computer networks is very wide and in comparison with the evolution of networks the development in this area is almost negligible. Therefore, only few tools can provide the mutual check of the pre-agreed conditions between the subscriber and the service provider.

That is the reason why the MONICA research group has decided to enter the research field of SLA. SLA Meter is our first application result of the research in the field of SLA.

II. Goals

The main goal of this paper is to provide a closer look into the design and development of the SLA Meter tool, which became a significant research subject of the MONICA research group in the field of SLA. Emphasis is also placed on the architecture of the metering tool.

III. Analysis and Design of the SLA Meter Tool

As mentioned above, one of the goals of the MONICA research group is the evaluation of the SLA. SLA is a document, which defines the relation between the subscriber and the service provider [1]. This document is really important for both sides. If it is used and treated in the right way, it can:

- identify and define the requirements of subscriber,
- provide a method for easier understanding,
- simplify complex problems,
- reduce conflict areas,
- promote the dialogue in case of conflicts,
- eliminate unrealistic expectations.

Yet, during the development other objectives had been preferred so far. The emphasis was placed on the establishment of a good basis for the successful solution of the other goals that can be found on the homepage of the research group [2].

ISBN 978-80-553-0890-6 ©2012 FEI TUKE

A. *The Architecture of the SLA Meter Tool*

Given the scope of the issue, it was decided to create a new application called SLA Meter. The basis for the origin of this application was not a market analysis, but the BasicMeter metering tool [3]. One of the main requirements was the modularity of the application for the possibility to easily expand and modify its parts.

The architecture of the SLA Meter tool [4] is described on Fig. 1. Components with blue colour are parts of the BasicMeter tool. These parts can be used in the first version of the SLA Meter without changes, but later they may be modified or even fully replaced. Everything will depend on the emerging needs and the environment of its deployment. Components with red colour represent the new parts of the SLA Meter tool. These parts are the Evaluator and the web interface. As one can see, it was decided to divide the uppermost layer into two parts. The reason for this decision was to separate the logical part from the graphical user interface. The Evaluator as the logical part of the SLA Meter calculates, analyses, evaluates and adjusts the records from the database so that the web interface as the part of the graphical user interface of the tool had only to display this information in an useful form for the user. The web interface also consists of a set of components with specific output format.

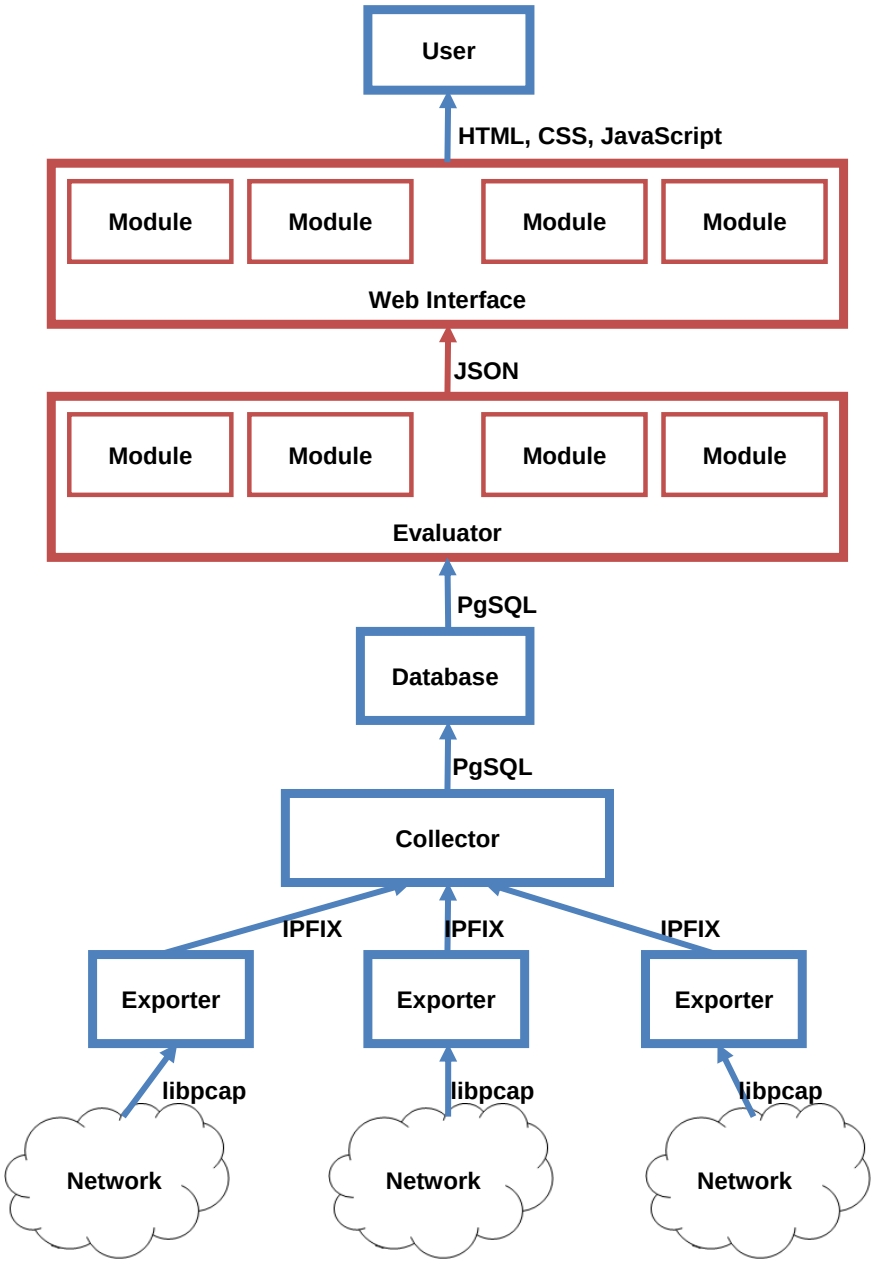


Fig. 1 The architecture of the SLA Meter tool.

B. The Exporting Process – BEEM

The exporter describes the lowest part of the BasicMeter tool [5]. The major and only developed application of this part of the tool is BEEM. It forms a connection of the tool with the monitored network, from which BEEM – according to its actual configuration by the libpcap library – is capturing packets from the network traffic. Information about these packets are subsequently transformed into IPFIX or NetFlow flow records and exported to the middle part of the BasicMeter tool.

C. The Collecting Process – JXColl

The collector represents the middle part of the BasicMeter architecture [6]. Currently, the only developed kind of the collector is JXColl. JXColl (Java XML Collector) is an application that provides the collection of IP Flow data received from various exporters (BEEMs). It also serves as a pre-processor of some data for their easier analysis. JXColl can be arbitrarily configured by its XML configuration file. It can store the collected data into a database, or can make them directly accessible by the means of the ACP protocol [7], [8].

D. The Evaluator of the SLA Meter Tool

The Evaluator is an application whose primary goal is to prepare the data for the web interface of the SLA Meter tool. The Evaluator – on the basis of some requirements – processes the IPFIX records from the database and creates statistical and analytical information about the nature of the monitored network traffic. Since the SLA Meter is planned for a wide application deployment, it is impossible to define its whole functionality during its design. For this reason its architecture is modular, i.e. it consists of a framework and a set of embeddable components. The components calculate the given data and provide them in a simple numeric form, from which the framework subsequently generates the responses.

E. Communication Principles Between the Evaluator and the Web Interface

It was decided to form the communication between the web interface and Evaluator of the SLA Meter tool on a JSON standard. JSON offers a structure, which can be easily read by a human and computers. The decisive advantage over the similar XML format is the file size, which is several times smaller, so it considerably reduces the amount of the transported data.

F. Web Interface of the SLA Meter Tool

The web interface is a web application whose function is to provide a graphical user interface for the communication between the user and the SLA Meter tool. Its architecture is modular. It consists of a framework and a set of embeddable components. The role of the components is to create the individual parts of the resulting web page in a form of HTML code fragments. These fragments – on the basis of a configuration file – are subsequently arranged into a final form by the framework.

IV. Implementation of the Evaluator Tool

The implementation of the Evaluator tool consists of the creation of the individual modules and the creation of the framework into which the modules will be inserted. For the implementation itself the Java programming language was chosen.

A. Implementation of the Framework

The main goal of the implementation process was to create a web service (EvalFramework), which includes the *getResult* method. This method is the only communication point of the Evaluator with the other applications of the SLA Meter tool. It was necessary to implement the received messages formatted as JSON. It was also required to divide this message between the individual modules. This was achieved by the use of json-1.0 library. First, the creation of the *JSONObject* object obtained from the received message was implemented. Subsequently, the field of JSONs was selected from the *modules* element. Each of these JSONs is a requirement for an evaluation module of the Evaluator. So it was necessary to transport the right requirement to the correct location. For this the *ModuleManager* class with its call method was created. This method is given the name of the module with the request, which is waiting for an answer. After receiving the responses from each of the required modules the response was completed and sent back to the web interface.

```
String response = "{\"modules\":[ ";
try {
  //creating a JSONObject from received json
  JSONObject jsonObject = new JSONObject(jsonText);
  //getting the array of modules
  JSONArray jsonArray = jsonObject.getJSONArray("modules");
  //getting through the json array
  for (int i = 0; i < jsonArray.length(); i++) {
    //getting a JSONObject from the array
    JSONObject JSON = jsonArray.getJSONObject(i);
    //executing of the method call
    String moduleResponse = moduleCall
                .call(json.getString("name"), json.toString());
    if (moduleResponse != null) {
      response += moduleResponse + ",";
    }
  }
  response = response.substring(0, response.length() - 1);
  //closing of the response message
  response += "]}";
```

For the implementation of the communication between the framework of the Evaluator and its evaluating components the *ModuleManager* class has been created. This class includes a list of all existing evaluating modules. These modules are stored as enumerate types. The class includes also all methods which enable to call the web method of each module. These methods are result of the registration of the framework Evaluator as a client of individual services of the evaluating modules. We can get these methods using the WSDL address of the web services.

But the main part of the *ModuleManager* class is the method *call*. This method was implemented to get all results of each module based on the request from the client of the Evaluator. The parameters are the name of the module and the request for the module formatted as JSON. After receiving the parameters the method gets through the module list. When the required module exists and it is allowed to use it, the method gets its communication method using reflection [9] and invokes it. The received data are the return value of the method.

```
for (Module m : Module.values()) {
  if (m.getName().equals(moduleName.toUpperCase())) {
    if (m.isAllowed()) {
      try {
        Method method = ModuleManager.class.getDeclaredMethod(
                        m.getMethod(), new Class[]{String.class});
        result = (String) method.invoke(this.getClass(), request);
```

B. Implementation of the Modules

To create a new module it is important to follow the agreed standards. The following template is to be used.

```
@WebService(serviceName = "NameOfTheService")
public class NameOfTheService {
@WebMethod(operationName = "getNameOfTheService")
public String getNameOfTheServiceResult
   (@WebParam(name = "request") String request){}}
```

V. Implementation of the Web Interface

The implementation of the web interface consists again of the implementation of the framework and modules that can be inserted there. The used language was Python.

A. Implementation of the Framework

The framework's main part is the *Distributor* class. The class is responsible for the communication with the Evaluator, with each module of the web interface and with the other parts of the framework.

To secure the communication with the Evaluator the *callWebService* method has been created. For the connection to the service the library *SOAPpy* has been used, in particular the *SOAPProxy* class and the *_ns* method.

```
def callWebService(self, json):
        url = 'http://localhost:8084/Evaluator/EvalFramework?wsdl'
        namespace = 'http://service/'
        server = SOAPProxy(url)
        return server._ns(namespace).getResult(jsonText = json)
```

For the communication between the framework and modules the reflection has been used by means of which individual modules of the web interface were called. To get each request of the module the *getJSONRequest* method has been created.

```
def getJSONRequest(self, blockList):
        json = "{\"modules\":[ "
        for block in blockList:
            instance = self.getClass(block.className)()
            name = block.name
            name = name[1:-1]
            jsonReq = instance.getRequest()
            if jsonReq != "":
                json = json + jsonReq + ","
        json = json[:-1]
        json = json + "]}"
        return json
```

For getting the HTML code of each module it was necessary to split the received responses from the Evaluator formatted as JSON. It was necessary to select the part of the response which contains the results of the request of each module. By successively getting through these responses the *getBlockHTML* method of the concrete module was executed using reflection.

```
data = json.loads(jsonRes)
jsonDataModules = data['modules']

for m in jsonDataModules:
       for block in blockList:
                name = block.name
                instance = self.getClass(block.className)()
                if str.upper(name) == m['name']:
                    global jsonResponse
                    jsonResponse = m
                    html = instance.getBlockHTML(m)
                    print html
                    block.setRenderedOutput(html)
```

For displaying the HTML code of each module the *views* class has been created. In this class the method for the processing and displaying of the web page content has been implemented. This page consists of parts created by each module.

```
def client(request):
    distributor = Distributor()
    blockList = distributor.getHTML(True)
    blockList = sorted(blockList, key=operator.attrgetter("weight"))
    activeRegions = []
    for block in blockList:
        activeRegions.append(block.region)
    return render_to_response('client.html', {
        'blocks': blockList,
        'activeRegions': activeRegions,
        }, context_instance=RequestContext(request))
```

B. Implementation of the Modules

For adding a new module into the framework it is necessary to create a new directory in the *apps/components* path of the web interface project for the module. The new module has to fulfil the definitions (has to contain the *getRequest* and *getBlockHTML* methods).

Then it is necessary to create an own HTML template and place it into the *templates* directory. The last step is to write the new module and the information about it into the configuration file of the web interface.

VI. Solution and Results

Based on the design we implemented the individual parts of the framework. The result is a dynamic web page displayed in Fig 2.

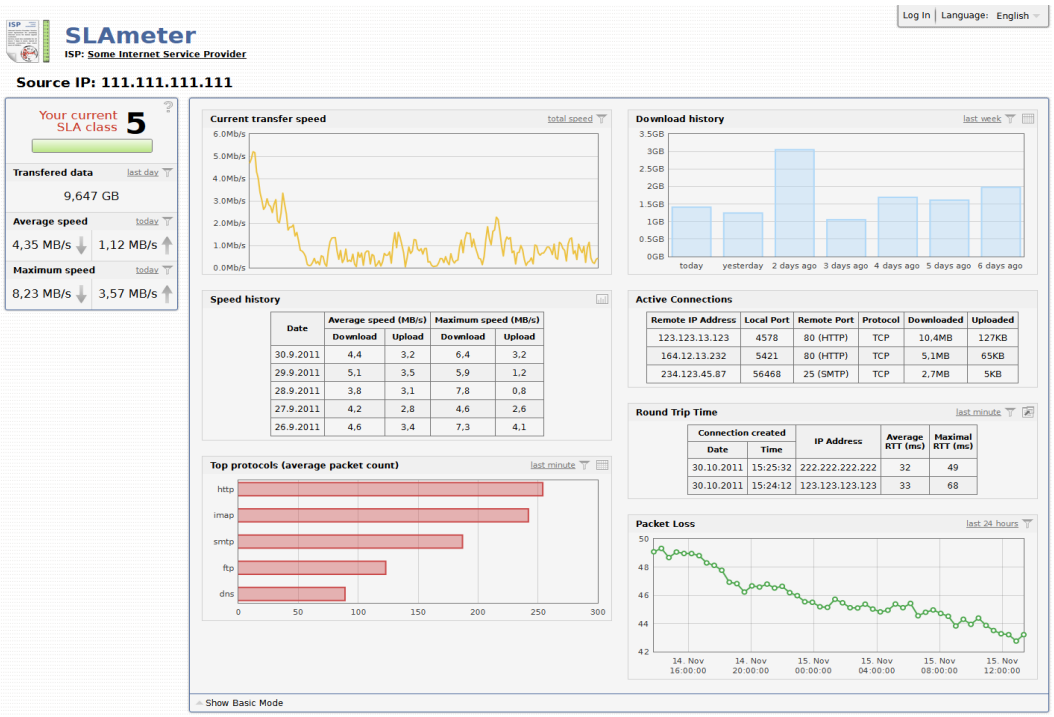


Fig. 2 Web interface of the SLA Meter tool.

VII. Acknowledgment

This work is the result of the project implementation: Development of the Center of Information and Communication Technologies for Knowledge Systems (ITMS project code: 26220120030) supported by the Research & Development Operational Program funded by the ERDF.

VIII. Conclusion

Within this contribution the SLA Meter tool has been described whose main function is to simplify the control of the pre-agreed conditions between the subscriber and service provider. The emphasis was on the architecture, which consists of the exporting, collecting and evaluating processes and also from the web interface, which is the user interface of the tool.

Since the requirements in the SLA area are developing dynamically, we have decided to design the modular architecture tool. The Evaluator and web interface consist of a framework and modules inserted into it, which ensure the required modularity. The use of web services and reflection for ensuring the modularity proved to be a very good solution. The advantage was the simple implementation, easy and reliable use.

Future work should be aimed at the development of new modules which would extend the tool's applicability in new domains, e.g. usage-based accounting or security analysis.

References

[1] G. Blokdijk, *Service Level Agreement 100 Success Secrets: SLA, Service Level Agreements, Service Level Management and Much More*. Newstead, Australia: Emereo Publishing, Jan. 2008.
[2] The MONICA Research Group, 2012. [Online]. Available: http://wiki.cnl.sk/Monica/WebHome
[3] F. Jakab, Ľ. Koščo, M. Potocký, J. Giertl, "Contribution to QoS Parameters Measurement: The BasicMeter Project," *International Conference on Emerging e-learning Technologies and Applications (ICETA)*, pp. 371–377, 2005.
[4] M. Antl, *Rámec vyhodnocovača a webového rozhrania nástroja SLA Meter*, Technická univerzita v Košiciach, Fakulta elektrotechniky a informatiky, 2012.
[5] T. Kecsey, *Optimalizácia meracieho a exportovacieho procesu nástroja BasicMeter*, Technická univerzita v Košiciach, Fakulta elektrotechniky a informatiky, 2012.
[6] A. Pekár, *Optimalizácia zhromažďovacieho procesu nástroja BasicMeter*, Technická univerzita v Košiciach, Fakulta elektrotechniky a informatiky, 2011.
[7] F. Jakab., J. Giertl, R. Jakab, M. Kaščák, "Improving Efficiency and Manageability in IPFIX Network Monitoring Platform," *International Network Conference (INC)*, pp. 81–88, July 2006.
[8] J. Giertl, R. Jakab, Ľ. Koščo, "Communication Protocol in Computer Network Performance Parameters Measurement," *International Information and Telecommunication Technologies Symposium (I2TS)*, pp. 161–162, Dec. 2005.
[9] I. R. Forman, *Java Reflection in Action*. Shelter Island, NY: Manning Publications, Oct. 2004.

The rôle of Abstract Data Types and Abstract Behavior Types

Marián POLÁK, Viliam SLODIČÁK

Department of Computers and Informatics, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Slovak Republic

marian.polak@student.tuke.sk, viliam.slodicak@tuke.sk

Abstract **– The notion of Abstract Data Types (ADT) has served as a foundation model for structured and object oriented programming.The most basic inherent property of an ADT is, that it provides a set of operations.**

The notion of Abstract Behavior Type (ABT) are a higher-level alternative to ADT and propose it as a proper foundation model for both components and their composition.An ABT defines an abstract behavior as a relation among a set of timed-data-streams, without specifying any detail about the operation.

Keywords **– Abstract Data Types, Abstract Behavior Types, Pop, Push, Enqueue, Dequeue**

I. INTRODUCTION

An ADT defines an algebra of operations with mathematically well-defined semantics, without specifying any detail about the implementation of those operations or the data structures they operate on to realize them. As such, ADT is a powerful abstraction and encapsulation mechanism that groups data together with their related operations into logically coherent and loosely-dependent entities, such as objects, yielding better structured programs. ADT has served as a foundation model for structured and object oriented programming for some thirty years.

We introduce the notion of ABT as a higher-level alternative to ADT and propose it as a proper foundation model for both components and their composition. The ABT model is not meant as a substitute for ADT. Instead, it is meant as a foundation for *programming with components* as opposed to *programming of components* which is what the ADT model and object oriented programming have proved to be suitable and efective for.

An ABT defines an abstract behavior as a relation among a set of
timed- data- streams, without specifying any detail about the operations that may be used to implement such behavior or the data types it may manipulate for its realization. In contrast with the algebraic underpinnings of the ADT model, the (generally) infinite streams that are the elements of behavior in the ABT model naturally lend themselves to the coalgebraic techniques and the coinduction reasoning principle that have recently been developed as a general theory to describe the behavior of dynamic systems. The ABT model supports a much looser coupling than is possible with ADT and is inherently amenable to exogenous coordination. [1]

II. GOALS

The goal of this work is to describe two abstract types, and they are ADT and ABT where ABT are a higher-level alternative to ADT.

III. ANALYSIS

A. Abstract Data Types (ADT)

Definiton 3.1. ADT consist of a specification of the possible values of the data type and a specification of the operations that can be performed on those values in termsof the operations inputs, outputs and effects.

We are all used to dealing with the primitive data types as abstract data types.It is quite likely that you don't know the details of how values of type *double* are represented as sequences of bits. The details are , in fact, rather complicated. However , you know how to work with *double* values by adding them, multiplying them , truncating them to values of type *int* , inputting and

outputting them, and so on. To work with values of type *double*, you only need to know how to use these operations and what they do. You don't need to know how they are implemented.

When you create your own new data types - stacks , queues , trees , and even more complicated structures - you are necessarily deeply involved in the data representation and in the implementation of the operations.

However , when you use these data types, you don't have to worry about the implementation. You want to be able to use them the way you use the built-in primitive types , as abstract data types. This is in conformance with one of the central principles of software engineering : *encapsulation* , also knows as information hoding , modularity , and separation of interface from implementation.

1) Stacks and Queues: The difference is the set of operations defined for the data type. In particular , the difference is in the meaning of the operation that removes one item from the data structure : A stack's *pop* operation removes the item that has been on the stack for the shortest amount of time, while a queue's *dequeue* operation removed the item that has been on the queue the longest. We define the ADT *"stack of type BaseType "* where *BaseType* could be integer , string , or any other data type.

A stack is a last in , first out (LIFO) data structure.The item that is in line to be the next one removed from the stack is the one that was added last , that is most recently.The operations on a stack include adding and removing items.

Definicion 3.2. For any data type, *BaseType* ,we define an abstract data type **Stack of BaseType** .The possible values of this type are sequence of items of type *BaseType* (including the sequence of lenght zero).
The operarions of the ADT are :

- **push(x) :** Adds an item , *x*, of type *BaseType* to the stack; that is , modifies the value of the stack by appending *x* to the sequence of item that is already on the stack. No return value
- **pop :** Removes an item from the stack and returns that item. The removed item is the one that was added most recently by the *push* operation. It is an error to apply this operation to an empty stack
- **makeEmpty :** Removes all items from the stack.No return value.
- **isEmpty :** Returns a value of type *boolean* , which is true if the stack contains no items and is false if there is at least one item on the stack.

We will use this definition of the stack ADT , but it is not the only possible definition.In some variations , for example , the *pop* operation removes an item from the stack but does not return it, and there is a separate *top* operation that tells you item on the top of the stack without removing it. Some implementation of stacks have a fixed maximum size. The definition of the *queue* ADT is very similar , with operations named *enqueue* and *dequeue* replacing *push* and *pop* :

Definition 3.3. For any data type, *BaseType*,we define an abstract data type **Queue of BaseType**. The possible values of this type are sequence of items of type *BaseType*(including the sequence of lenght zero).
The operarions of the ADT are :

- **enqueue(x) :** Adds an item , *x*, of type *BaseType* to the queue, that is , modifies the value of the queue by appending x to the sequence of item that is already on the queue. No return value.
- **dequeue :** Removes an item from the queue and returns that item. The item removed is the one that has been in the queue the longest.It is an error to apply this operation to an empty queue.
- **makeEmpty :** Removes all items from the queue.No return value.
- **isEmpty :** Returns a value of type *boolean* , which is true if the queue contains no items and is false if there is at least one item in the queue.

A queue is a first in , first out (FIFO) data structure. The item that is next in line is the one that arrived first in the queue and that has been waiting in the queue the longest.[2]

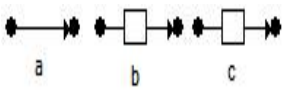


Fig. 1 Basic channels in ABT



Fig. 2 Asynchronous FIFO channel contains the data item D (a) and Synchronous drain channel (b)

2) Properties of abstract data type: : The most important properties of abstract data type are :

- **Generality of the implementation :** once designed ADT can be built and used without problems in any program
- **A precise description :** link between implementation and interface must be clear and complete.
- **Simplicity :** in use , the user need not worry about the internal implementation and administration of ADT in mind.
- **Encapsulation :** interface should be submitted as the closed part.The user should know exactly what ADT does , but not how it works
- **Integrity :** The user can not interfere in the internal data structures.This greatly reduces the risk of unwanted deletion or modification of already stored data.
- **Modularity :** principle of "modular" programming is transparent and allows easy exchange of part of code. When searching for error individual modules can be considered as compact units. By improving the ADT it is not necessary to interfene into all program.[3]

B. Abstract Behavioral Typed (ABT)

An ABT defines an abstract behavior as a relation among a set of *timed- data- streams*, without specifying any detail about:

- the operations that may be used to implement such behavior
- or the data types it may manipulate for its realization.

A timed-data-stream is a twin pair of streams $\langle \alpha, a \rangle$ consisting of a data stream α and a time stream a with the interpretation that $\forall_i \geq 0$, the data item $\alpha(i)$ appears at its corresponding time moment a(i).Used to model the data flow through channel ends.

1) ABT Examples:

A. **Basic channels**
   - *Synchronous channel* as in Figure 1.a. : $\langle \alpha, a \rangle \, Sync \, \langle \beta, b \rangle \equiv \langle \alpha, a \rangle = \langle \beta, b \rangle$
   - *Asynchronous unbounded FIFO channel* as in Figure 1.b. : $\langle \alpha, a \rangle \, FIFO \, \langle \beta, b \rangle \equiv \alpha = \beta \wedge a < b$
   - *Asynchronous FIFO channel with the bounded capacity of 1* as in Figure 1.c. : $\langle \alpha, a \rangle \, FIFO_1 \, \langle \beta, b \rangle \equiv \alpha = \beta \wedge a < b < a'$

B. **Asynchronous FIFO channel with the bounded capacity of 1 and initially contain the data item D** as in Figure 2.a. : $\langle \alpha, a \rangle \, FIFO_1(D) \, \langle \beta, b \rangle \equiv \beta(0) = \beta' \wedge b < a < b'$

C. **Synchronous drain channel** as in Figure 2.b. :
   - Two source ends, cannot produce any data items.
   - Every data item is lost

   $SyncDrain \, (\langle \alpha, a \rangle, \langle \beta, b \rangle ;) \equiv a = b$ [4]

2) Merge and replication:

A. *The merger ABT* is defined as:
$M(\langle \alpha, a \rangle, \langle \beta, b \rangle, \langle \gamma, c \rangle) \equiv a(0) \neq b(0) \wedge \quad where$
$\alpha(0) = \gamma(0) \wedge a(0) = c(0) \wedge M(\langle \alpha', a' \rangle, \langle \beta, b \rangle ; \langle \gamma', c' \rangle) \quad if \quad a(0) < b(0) \quad and$
$\beta(0) = \gamma(0) \wedge b(0) = c(0) \wedge M(\langle \alpha, a \rangle, \langle \beta', b' \rangle ; \langle \gamma', c' \rangle) \quad otherwise$

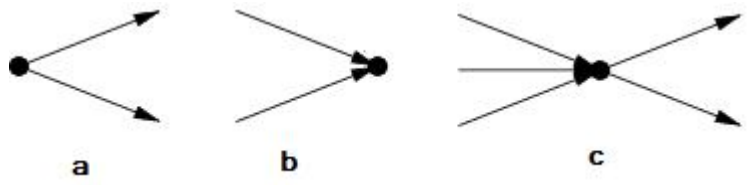


Fig. 3 Types of nodes in Reo.

B. *The replicator ABT* is defined as:
$Rpl\left(\langle\alpha, a\rangle; \langle\beta, b\rangle, \langle\gamma, c\rangle\right) \equiv \beta = \alpha \wedge \gamma = a \wedge b = a \wedge c = a$ [4]

3) Reo: The ABT model is too low level to provide non-trivial coordination , an effective exogenouns coordination model is needed.Reo is the channel-based exogenous coordination model wherein complex coordinators, called connectors are compositionally buit of simple ones.

- The simplest connectors in Reo are channels.
- A connector is a set of channel ends and treir connecting channels organized in a graph of nodes and edges (channels) such that
  - Zero or more channel ends coincide on every node.
  - Every channel end coincides on exactly one node.
  - There is an edge between two nodes if and only if there is a channel one end of which conincind on each of those nodes.
- Powerful mechanism for connector conduction : channel composition [4]

4) Reo defines three types of nodes :

A. *Source nodes* as in Figure 3.a. : $Src\left(N\right) \neq \emptyset \wedge Snk\left(N\right) = \emptyset$
   - Src(N) and Snk(N) : sets of source and sink channel ends coincide on N, respectively.
   - A source node replicated every data item written to it as soon as all of its coincident source channel ends can consume that data item.

B. *Sink nodes* as in Figure 3.b.: $Src\left(N\right) = \emptyset \wedge Snk\left(N\right) \neq \emptyset$
   - non-deterministically selects one of the data items available through its coincident sink channel ends.

C. *Mixed nodes* as in Figure 3.c.: $Src\left(N\right) \neq \emptyset \wedge Snk\left(N\right) \neq \emptyset$
   - Combines the behavior of a sink node and a source node.
   - Every time a data item is available through its coincident sink channel ends it non-determinictically selects an appropriate data item and replicated that data item into all of its coincident source channel ends. [4]

IV. Conclusion

The goal of this work was to describe ADT and ABT.It has been described what it is ADT and ABT with examples and possibly even pictures and describe the properties. In the following article can be ADT and ABT can be compared against each other.

V. Acknowledgment

This work has been supported by KEGA Grant No. 050TUKE-4/2012:Application of Virtual Reality Technologies in Teaching Formal Methods.

References

[1] Farhad Arbab: Abstract Behavior Types: A Foundation Model for Components and Their Composition, Centre for Mathematics and Computer Science(CWI),Kruislaan 413,1098 SJ Amsterdam,The Netherlands, 2 p.
[2] David J.Eck : Data Structure and Algorithms,Department of Mathematics and Computer Science Hobart and William Smith Colleges,Geneva, New York, USA ,2004 , pp. 11–13
[3] kiwiki,Data type,online url address: kiwiki.fmtnuni.sk/index.php/Dtov_typ
[4] Nadia Ramjiawan: Seminar on Component-based Software Engineering ,Component Coordination , LIACS - Leiden University ,USA , 2005, pp. 4–9

3D Training Manual with Augmented Reality Technologies

Monika RAŠKOVSKÁ, Branislav SOBOTA, Martin VARGA

Department of Computers and Informatics, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Slovak Republic

monika.raskovska@gmail.com, branislav.sobota@tuke.sk, martin.varga@tuke.sk

***Abstract* — This work deals with creating augmented reality application with indirect view. The main ambition is to develop software environment for 3D manual in programming languages C and C++ based on tool ARToolKit. Application will support VRML format and will be managed script written in Lua. Finally, it describes needs and limitations of creating virtual models and their modeling for blending 3D manual form.**

***Keywords* — ARToolKit, augmented reality, Lua, virtual reality, VRML**

I. INTRODUCTION

Virtual reality technologies are today the latest ways of communication human-computer. Quickly began to create the applications that this technology used in varying degrees. After the success this applications in the world of computer games, virtual reality today gets into the sphere of training pilots, drivers of medicine and surgery, the industry in a manuals, but also in education and research. Due to the high deployment and development of technology, is possible that virtual reality will become a part of computer equipment in the nearly future [1].

Different types of statistics have concluded that the study at the basis of passing information from textbooks is not very effective form of learning. The best method of teaching is to obtain information based on own experiences. Student the discussing problem real sees, hears and realizes. Over the years textual forms of documents switched to electronic forms and user guides have received a new form the electronic manual. With the using the three-dimensional technology it comes in being a new type of e-manual and this was named 3D manual.

The topic of this work is to give some indication on how to develop augmented reality application and managed it with script [2]. Moreover, the work deals with creating 3D training manual and modeling the most optimal virtual objects for it. The thematic fixation of manual was chosen like building computer based on its hardware components need for system functioning.

II. ANALYSIS

A. Augmented reality

3D is space defined in three dimensions. Compared to two-dimensional space is defined on a new dimension of depth [3]. Virtual reality is defined as computer-generated, interactive three-dimensional environment, into which the user is embedded. The user sees a scene with each eye from a different angle, and this is the way to create a credible 3D scene. By Paul Milgram belongs to the virtual reality technologies the concept of Mixed reality [4] (see on Fig. 1).

Augmented reality means that the real world is enriched by virtual elements. Moreover concept Augmented virtuality means that the virtual world is enriched by elements of real word. Mixed reality covers both concepts of augmented reality and augmented virtuality. It is actually a continuous space between reality and virtuality and is bounded on one side real and on the other side virtual environment [4].

Based on how a user sees mixed reality there can be two types of systems:

- *Optical see-through (direct) systems* where the user sees real world directly and computer generated objects are added to this view. This category of systems usually works with semi-transparent displays.

- *Video see-through (indirect) systems* where captured real world image with added virtual objects is displayed to the user. This is usually realized via camera – display system [4].

According to the method how virtual objects are aligned with real scene image there are two systems use:

- *Marker systems* – special markers are used in a real scene. The markers are then recognized during runtime and replaced with virtual objects.
- *Markerless systems* – processing and inserting of virtual objects is without special markers. Additional information is needed, for example image (e.g. photo (semi-markerless system)), face recognition, GPS data, etc.

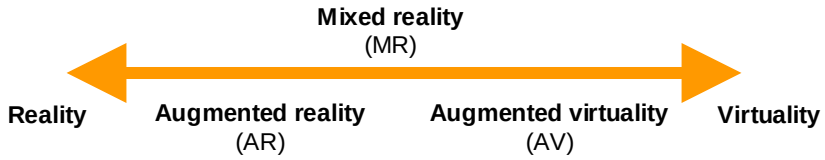


Fig. 1 Milgram's definition of real to virtual world transition

B. *Developmental technologies*

The one of the most integrated tools for developing markered augmented reality applications is ARToolKit [5] (Software library for building AR applications created by Human Interface Technology laboratory: *http://www.hitl.washington.edu/ artoolkit/*). It is a collection of libraries written in programming languages C and C++. ARToolKit allows creating augmented reality application based on camera and markers. Principle of ARToolKit applications is on the next figure Fig. 2.

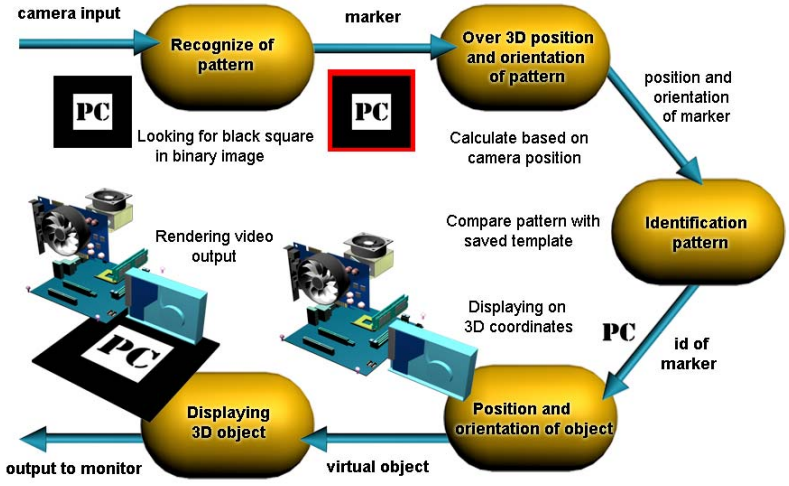


Fig. 2 Principle of ARToolKit application

The camera reads in the real world and in each frame from video capture is looking for a marker. Marker is a black square with some pattern in the middle. The software detect marker and calculate the position of marker. Based on this position is created transformation matrix, in which virtual elements are displayed later. ARToolKit allows working with multiple markers and cameras. Enables in the real word recognize marker and on its position display elements of virtual world.

VRML [6] is a text format that describes 3D scene much more complex than relatively simple *.obj* files. In this format are data written in a similar way as a programming language. In addition of the definition 3D object contains information about lighting scene, sounds and transformations of virtual object. VRML files are often called "worlds" and have *.wrl* suffix.

Lua [7] is a modern high-level dynamic typed scripting language that has syntax inspired by languages Pascal and Modula. Provides an easy work and has inbuilt compiler. It is implemented as a library written in ANSI C. With these option is Lua nowadays mostly used in game development and it is possible to build Lua scripts in another application. Also other

scripting languages as Ruby [8] or Python [9] are usable in this case.

III. Solution and Results

Created application makes for environment of manual. It was created as augmented reality application based on tool ARToolKit. This tool offers a few basic applications. Based on this basic tool has been created software for manual. It was employed in recognizing markers and it has been added on application and opportunity working with multiple patterns. Markers so are used for displaying virtual models and it belongs to each pattern another virtual models. By the help of one marker it is possible switching between individual manual's steps. The main logic of manual is commanded by script. Its commission is to set up index of virtual model, that about to be actual for displaying. Application workflow so been divided into two phases. First uses multiple patterns and characterizes the imaginary computer building. Virtual models replace the real components and they are instead used for engaging. The following figure (Fig. 3) shows this case.

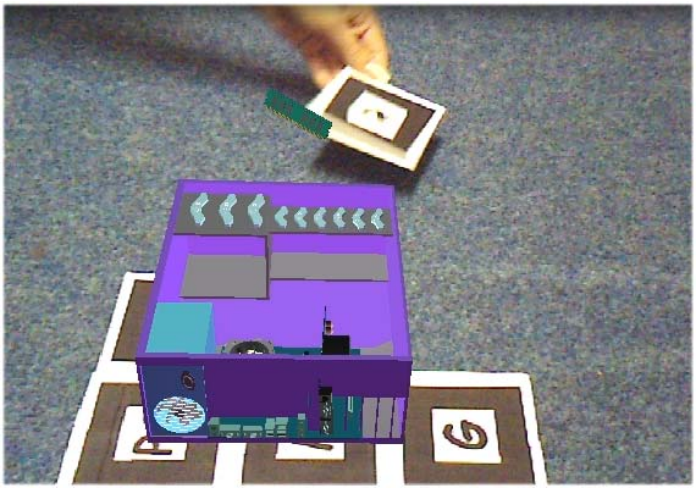


Fig. 3 First phase of maual

Second phase uses one pattern for displaying multiple virtual objects. Each virtual model is shown separately and it can be controlled by keyboard events. This phase characterized graphically display the involvement of real hardware computer components. Here virtual models with animation represent how to involve the real components. The following figure (Fig. 4) shows this case.

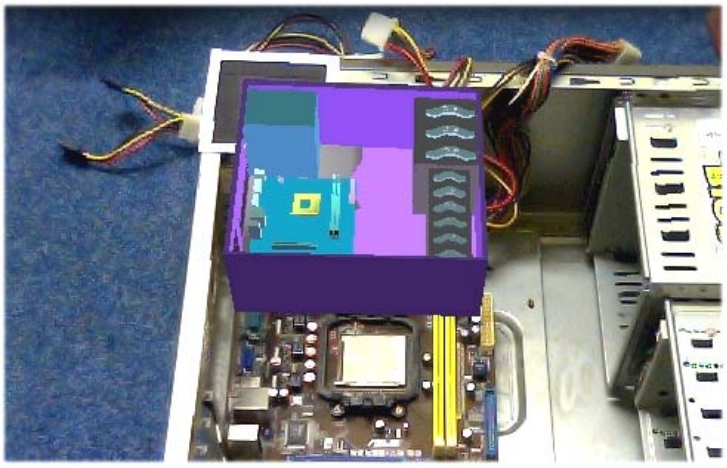


Fig. 4 Second phase of maual

Therefore the Application 3D-manual allows work with the real but also with virtual computer components and it tests their proper integration. If the integration is right or no, it is provided by recognized marker position. The application has been extended to support play various sound recordings for its improved functionality.

Application can be used for different purposes by changing virtual models and possible management apps. On next figure is shown the example of application employment for the establishment of production lines from its individual modules (Fig. 5.).

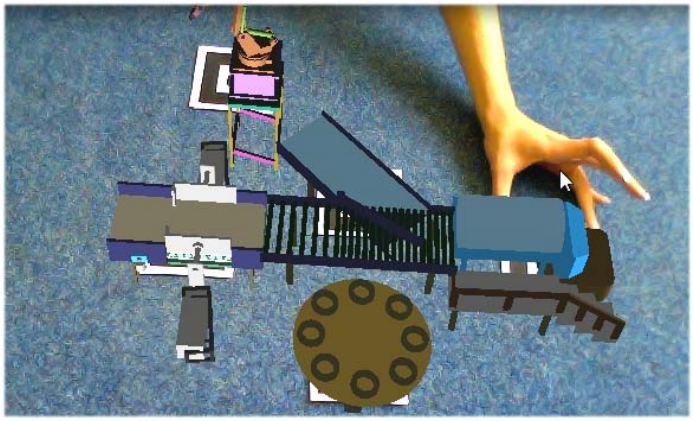


Fig. 5 Alternative to use application

IV. CONCLUSION

This paper presents the programming environment for 3D manual form using mixed/augmented reality technology. Theme of manual was chosen as the involvement of computer hardware components, or its individual parts.

For creating manual was used application with indirect view, programming languages C and C++ and tool ARToolKit. Script to managed application was written in language Lua. Application 3D-manual works with one camera and with multiple markers. Virtual models for blending manual was created in environment 3D Studio Max, Blender or in Google SketchUp.

Application work with lot of types of patterns, with many different virtual objects and based on these facts is this application universal to use it in many spheres of problems. Considering today's technology and capabilities is a solution of application optimal and actual.

ACKNOWLEDGMENT

This work is the result of the project implementation: Development of the Center of Information and Communication Technologies for Knowledge Systems (ITMS project code: 26220120030) supported by the Research & Development Operational Program funded by the ERDF.

REFERENCES

[1] F. Hrozek, B. Sobota, Cs. Szabó, Š. Korečko, M. Varga, P. Ivančák," Augmented reality application in parallel computing system", 7th International Workshop on Grid Computing for Complex Problems, Bratislava, Slovakia, 24 - 26 October 2011, Bratislava, Slovakia, Institute of Informatics Slovak Academy of Science, 2011, pp. 118-125, 978-80-970145-5-1

[2] B. Sobota, R. Janošo,"3D interface Based On Augmented Reality In Client Server Environment", Journal of information, control and management systems, Vol. 8, No. 3, 2010, pp. 247-256, ISSN 1336-1716 .

[3] Azuma R., "Tracking Requirements for Augmented Reality", Communications of the ACM Vol. 36, No. 7, 1993, pp. 50-51.

[4] P. Milgram; F. Kishino, "A Taxonomy of Mixed Reality Visual Displays", IEICE Transactions on Information Systems, Vol. E77-D, No. 12, 1994, pp. 1321-1329.

[5] ARToolKit Documentation, Human Interface Technology Laboratory, url: http://www.hitl.washington.edu/artoolkit/documentation/

[6] VRML – Virtual Reality Modeling Language - http://www.w3.org/MarkUp/VRML/

[7] LUA - the programming language – official website, url: http://www.lua.org/

[8] Ruby official website, url: http://www.ruby-lang.org/en/

[9] Python Programming Language – official website, url: http://www.python.org/

Geographical Information Systems and 3D Visualization Technologies

Branislav SOBOTA, František HROZEK, Lukáš FEDORKO

Department of Computers and Informatics, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Slovak Republic

branislav.sobota@tuke.sk, frantisek.hrozek@tuke.sk, lukas.fedorko@student.tuke.sk

***Abstract* — The aim of this paper are geographical information systems (GIS) and 3D technologies that are used for their visualization. Paper is divided into three parts. First part presents maps and standards that create core of every GIS. Technologies and tools for 3D visualization are content of the second part. Presented 3D modeling tool in this part is Google SketchUp. Third part presents two GIS created at Department of Computers and Informatics at Technical University of Košice.**

***Keywords* — geographical information systems, computer graphics, 3D modeling, visualization**

I. INTRODUCTION

An *information system* (IS) is a collection of people, methods, and technological resources providing data collection, transfer, distribution, retaining and processing. IS aim to provide a systematically ordered collection of information to end-users or to other systems (e. g. controllers) depending on the requests received.

A *geographical information system* [1] (GIS) is an IS designed to work with data that represent geographical or spatial coordinates. These data can be further analyzed and tatistically evaluated. Another way to specify a GIS is that it is a specialized IS defined as collection of computer hardware and software and geographical data designed for effective gathering, retaining, editing, processing, analysis and visualization of all kinds of geographical information. Generally, it is a spatial specialized IS.

This paper aims on the 3D visualization technologies of GIS. First part presents maps and standards that are used in GIS and creates core of every GIS. Second part deals with the problematic of technologies and tools for 3D visualization of GIS. Third part presents two GIS that were created at Department of Computers and Informatics at Technical University of Košice (DCI TUKE).

II. CORE OF THE GIS - MAPS AND STANDARDS

A. Maps

The core of a GIS is the map. Most common internal representations of these maps use vector graphics. This representation supports fast changes because of the numerical values of coordinates. Coordinates inside the map include the dimensions of the terrain objects, therefore the process of creation of such a map starts with terrain object measuring. The resolution (e. g. measurement unit) depends on the aims of use. Another way to get the required information is to use existing data from cadastral maps. In the case of implementing changes, the preparation of inputs for a GIS is called geometric plan. It includes information about the state of the map *before* and *after* the changes takes effect (all changing attributes). Geometric plan preparation consists of the following activities:

- Collecting actual data from the IS: graphical, geodetic and text information.
- Measuring selected terrain objects – it depends on the difficulty of the process and terrain obstacles.
- Processing and documenting the gathered data: new and changed coordinates, dimensions, non-graphical information.

ISBN 978-80-553-0890-6 © 2012 FEI TUKE

B. Standards

Introducing standardized solutions and formats into industry results to compatibility and interoperability of large systems dealing with spatial and geographical information that a GIS may consist of.

The basic standards on geodetics and GIS in Europe are the European Terrestrial Reference System (ETRS89) and the European Vertical Reference System (EVRS2000). Since spring 2000, there is the European Spatial Reference System (ESRS) as the union of the two above-mentioned systems.

III. Existing Technologies and Tools

First, the process of creating a GIS requires *universal software* tools, such as operating system, database management system and graphical design tool (e. g. CAD for creating graphical models, Autodesk's AutoCAD or MicroStation from Bentley Inc.). Second, we need special *geoinformation software,* such as programs for automatic terrain digitalization (e.g. TerraModeller, Descartes) or programs for processing information about engineering networks (e.g. Intergraph's FRAMME). Many GIS use a database management system (DBMS) to keep data organized. Some use it only to store attribute data, and the processing of them and other data is done by a separate system, these GIS are called *dual GIS*. Other ones store all data in a database and use its services to process these data in a certain way. This type we call *integrated GIS*. OpenGIS [2] is an initiative about getting dual GIS into a distributed computing environment using its abilities to use external relational DBMS and making it open source.

A. Modeling using Polygons

Polygonal modeling is a term of computer graphics describing the method of using polygons as spatial object representation, e. g. we represent each object as a collection of the polygons of the surface [3]. This technique sounds suitable for real-time graphics and for GIS as well due to the line-by-line rendering mechanism of visualization. Polygonal model of University library is shown in Fig. 1.

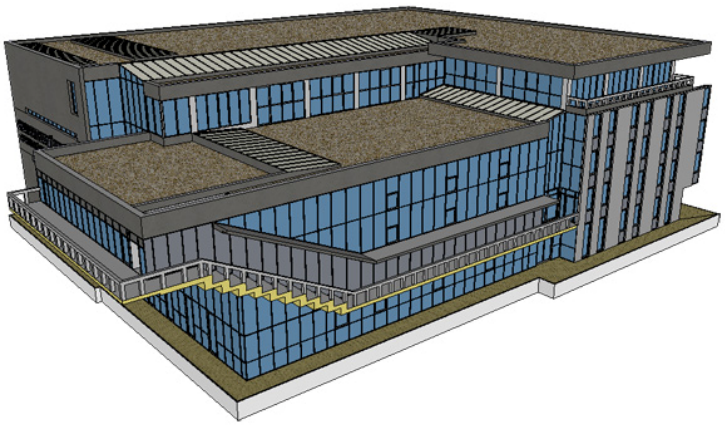


Fig. 1 Polygonal model of University library

B. Laser Scanning

Laser scanning allows effective scanning of terrain without direct contact to it. This technique is suitable for gathering information about the existing state of objects of size medium and above. Scanning is fast and produces high quality data that can be further processed (e.g. building 3D models of objects and terrain) and visualized.

The results of the scan are milliards of points called *points cloud* [4] [5], which are understandable by some CAD systems for digitalization and 3D modeling. *Points cloud* of Štátne divadlo Košice is shown in Fig. 2.



Fig. 2 Points cloud of Štátne divadlo Košice

C. 3D Modeling and Visualization

With laser scanning or even other form of scanning of towns form the air, we get photos (scans) with a very good resolution that can be transformed to 3D urban area model. These kinds of models are even more required by new generation of GIS.

One of the benefits of this modeling is the option to extend any element by additional not necessarily graphical attributes, e. g. we can attach to a larger dwelling complex the database of all owners, lodgers etc.

After the model is complete, it comes to visualization [6]. We say, visualization is analysis and presentation of spatial relations between the modeled 3D objects. Visualization might be of different forms, but the most common ones are passages between the modeled objects or flights above them. These could be animations only or interactive presentations.

We distinguish between three types of 3D models of urban areas: *Block model*, *Urban model* and *Detailed model*. Another method is to model the whole area including roads, vegetation etc., see Fig. 3, it is called *complex 3D model*.



Fig. 3 Complex 3D model

D. SketchUp Software Tool

As mentioned above, there are a lot of tools in this area. A very interesting one of them is Google's SketchUp [7] available commercially and/or for free (with less functions). Both versions support 3D modeling and uploading the created 3D objects into Google Earth and Google 3D Warehouse. SketchUp is very intuitive and easy to learn software to create virtual reality scenarios; it supports object modeling (intelligent cursor, push-pull, extrude for 3D drawing), camera positioning, coloring and texturing. Model of University library, shown in Fig. 1, was created in SketchUp.

IV. Examples of GIS That Use 3D Visualization

This section presents two GIS created at DCI. First one use panoramic pictures for information presentation and second one is based on PHP and use Python and 3D visualization core.

A. GIS with panoramic pictures

Work on this GIS was divided into tree parts: *modeling*, *presentation* and *visualization*. Visualization was provided by projecting of panoramic image (Fig. 4) on n-angle around camera. Position of n-angle and camera is constant. Visual effect of the movement is generated by changing of panoramic pictures. Actual projected panoramic image depends on actual viewer position on the map. The advantage of this approach is that it combines 2D approach (panoramic pictures) with 3D (panoramic pictures was created from photorealistic visualized model of city agglomeration). This solution supports low hardware requirements and high information value.

This GIS presents information about areal of Technical University of Kosice bounded by the adjacent streets: Letná, Komenského, Watsonová and Boženy Němcovej.



Fig. 4 One of the panoramic pictures used in GIS

B. Web Based GIS

This GIS is based on PHP and use Python and 3D visualization core (created at DCI). Python scripting language is used for creation of trajectories in 3D visualization core. 3D model was created in Google SketchUp. Fig. 5 (left) show photography of campus and Fig. 5. (right) shows created 3D model. Advantage of this GIS is its online accessibility.

This GIS presents information about campus at Jedlíkova 9, Košice.



Fig. 5 Photography of campus (left) and 3D model (right)

V. Conclusion

With emerging computer hardware capabilities in field of computer graphics and networks arise the possibilities of geographical information systems as well. Higher connection speeds and operational memory capacity together with better processor performance allow running applications with higher system and hardware requirements. GIS are one of these systems, especially when using 3D graphics as presentation layer.

In this paper was presented technologies and tools that are used for GIS creation. Also was presented approaches that can be used for 3D model creation. These approaches were: polygonal modeling and 3D scanning. For polygonal model creation was presented 3D modeling tool called Google SketchUp. Also two GIS, created at DCI TUKE, were presented in this paper.

ACKNOWLEDGMENT

This work is the result of the project implementation: Development of the Center of Information and Communication Technologies for Knowledge Systems (ITMS project code: 26220120030) supported by the Research & Development Operational Program funded by the ERDF.

REFERENCES

[1] J. Tuček, "Geografické informační systémy". Computer Press, Praha, 1998
[2] OpenGIS homepage. [online] URL: http://www.opengeospatial.org/ as of 10.5.2012
[3] B. Sobota, J. Perháč, I. Petz: *Surface modelling in 3D city information system, Journal of Computer Science and Control Systems*, 2, 2, 2009, pp. 53-56, ISSN 1844-6043
[4] Sobota Branislav, Rovňák Maroš, Szabó Csaba: 3D scanner data processing, Journal of Information, Control and Management Systems, Volume 7, 2, 2009, pp. 191-198, ISSN 1336-1716
[5] Hrozek František, Sobota Branislav, Szabó Csaba: Preservation of Historical Buildings Using Virtual Reality Technologies, Proceedings of the Eleventh International Conference on Informatics INFORMATICS'2011, Rožňava, Slovakia, November 16-18, 2011, Košice, Slovakia, EQUILIBRIA, s.r.o., 2011, pp. 164-168, ISBN 978-80-89284-94-8
[6] Sobota Branislav, Hrozek František, Repická Iveta: 3D Buildings Modeling and Visualization, Electrical Engineering and Informatics II, Proceeding of the Faculty of Electrical Engineering and Informatics of the Technical University of Košice, Košice, March 2011, Košice, FEI TU Košice, 2011, II, pp. 251-254, ISBN 978-80-553-0611-7
[7] Google SketchUp homepage. [online] URL: http://sketchup.google.com/ as of 10.5.2012

An Instruction Decoder and Disassembler Generator for EmuStudio Platform

Matúš SULÍR, Slavomír ŠIMOŇÁK

Department of Computers and Informatics, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Slovak Republic

matus.sulir@student.tuke.sk, slavomir.simonak@tuke.sk

Abstract **– In this paper, we present Edigen — a versatile generator of software-based instruction decoders and disassemblers compatible with emuStudio, a modular computer emulation platform. Edigen processes a file written in its own domain-specific language, containing the description of binary and textual instruction formats and generates Java source files. During the generation process, several tree transformations are performed.**

Keywords **– emulation, disassembler, instruction decoder, source code generation**

I. Introduction

EmuStudio [1] is a universal plugin-based emulation software, which means it is not oriented towards emulation of one specific platform, but rather a collection of various plugins tied together to form an emulated computer. Each plugin implements interfaces or extends classes located in the emuLib library. There exist plugins for processors, main memory, peripheral devices and compilers (usually assemblers). We will focus on one of the mandatory components of every emulated computer system — a central processing unit (CPU).

From the point of emulation in the emuStudio platform, there are three main tasks which the processor plugin must accomplish:

- read instructions from the main memory and decode them
- execute the decoded instructions
- provide a graphical user interface (GUI) for debugging

By the debugging user interface we mean a window or a set of windows displaying the processor state: the content of registers and a list of disassembled instructions around the current memory position.

II. Motivation

Originally, emuStudio platform did not separate the instruction decoding and execution processes. These two actions were mixed together within one class of the processor plugin. In addition, the information used during the decoding process were duplicated — once in the CPU itself and once in the disassembler GUI. Code duplication leads to severe problems [2].

Each instruction has its binary format, textual representation (mnemonic) and semantics. From the programmer's point of view, the most interesting task is the execution of the instruction, i.e. implementing its semantics. Writing an instruction decoder and disassembler can be considered labor and it usually does not require too much creativity. We have decided to create a domain-specific language and a source code generator for instruction decoders and disassemblers to make creation of processor plugins faster, more reliable and pleasurable. This is analogous to a compiler design. Writing the parser, which is traditionally a very repetitive task, is often left to a parser generator.

III. Specification Language

Edigen introduces its own domain-specific language for the specification of CPU instructions. An input file is divided into two parts:

1) binary format specification, from which the instruction decoder will be generated
2) textual format specification — a source of information for the disassembler generation

The LL grammar of the language follows.

$$
\begin{aligned}
Start &\longrightarrow Decoder\ \texttt{\%\%}\ Disassembler \\
Decoder &\longrightarrow Rule\ \{Rule\} \\
Rule &\longrightarrow RuleNameSet\ \texttt{=}\ VariantSet\ \texttt{;} \\
RuleNameSet &\longrightarrow RuleName\ \{\texttt{,}\ RuleName\} \\
RuleName &\longrightarrow \mathbf{id}\langle ident\rangle \\
VariantSet &\longrightarrow Variant\ \{\texttt{|}\ Variant\} \\
Variant &\longrightarrow [Return\ \texttt{:}]\ Pattern\{Pattern\} \\
Return &\longrightarrow \mathbf{id}\langle ident\rangle \mid \mathbf{string}\langle str\rangle \\
Pattern &\longrightarrow Subrule \mid Constant \\
SubRule &\longrightarrow SubRuleName\ [SubRuleLength] \\
SubRuleName &\longrightarrow \mathbf{id}\langle ident\rangle \\
SubRuleLength &\longrightarrow \mathbf{dec_number}\langle value\rangle \\
Constant &\longrightarrow \mathbf{hex_number}\langle value\rangle \mid \mathbf{bin_number}\langle value\rangle \\
Disassembler &\longrightarrow Format\ \{Format\} \\
Format &\longrightarrow FormatString\ \texttt{=}\ ValueSet\ \texttt{;} \\
FormatString &\longrightarrow \mathbf{string}\langle str\rangle \\
ValueSet &\longrightarrow Value\ \{Value\} \\
Value &\longrightarrow \mathbf{id}\langle ident\rangle
\end{aligned}
$$

We used JavaCC as the parser generator to produce the abstract syntax tree (AST) from the augmented grammar.

This is a simple specification file for a part of an imaginary processor:

```
instruction = "add": 0xFF12 dst(4) imm |
              "sub": 0xFF13 dst(4) src;
dst, src = "a": 0001 |
           "b": 0010;
imm  = i: i(4);

%%

"%s %s, %s" = instruction dst src;
"%s %s, %d" = instruction dst imm;
```

IV. Instruction Decoder

The aim of a generated instruction decoder is, given a byte or a list of bytes, to produce a map — which is a set of key–value pairs.

Instruction decoder specification consists of a set of rules. Each rule is identified by a name (or a set of names) and represents a possible key of the resulting map. A rule contains one or more variants. A variant consists of a recognition pattern and optionally a return value.

The instruction recognition starts at the first rule. During the decoding process, an input is matched against the recognition patterns. For each rule, only one variant can match the specific input. An input matches the variant if all constant bits of the variant (binary/hexadecimal digits in the example) are equal to the bits of the input at the corresponding positions.

If the selected variant returns a value, it is stored in the resulting map, using the rule identifier as a key. The value can be either a string or a sequence of bits.

The process continues hierarchically with all sub-rules of the variant until the whole instruction is decoded.

For the mentioned decoder example, given the input (in hexadecimal) “FF 12 10”, the output would be the map: {*instruction: “add”, dst: “a”, imm: 0000*}.

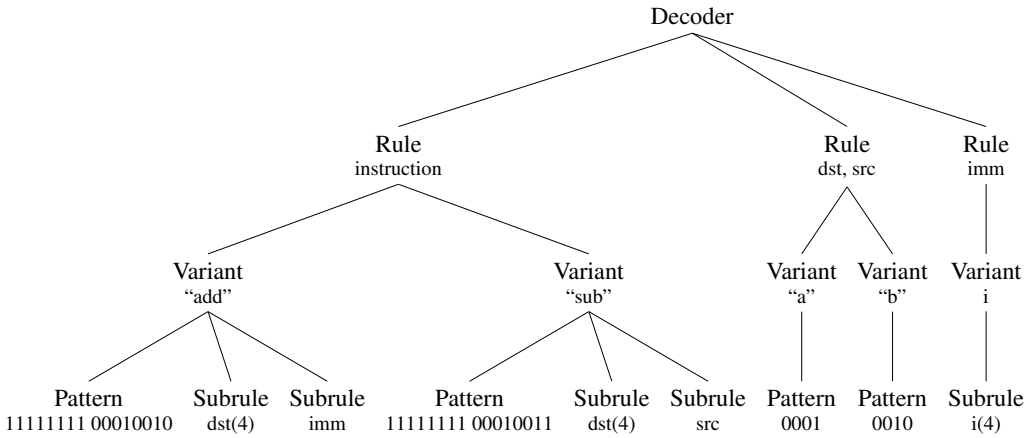


Fig. 1 An example of an initial decoder AST

V. Disassembler

The generated disassembler produces a mnemonic from a decoded instruction. At first, a format containing exactly the same set of keys as the map representing the decoded instruction is selected. The format values are substituted with the values contained in the map. Finally, the sequences of bits are transformed into a string representation using a format specifier, e.g. `"%x"` transforms the number 0 stored in two's complement into a hexadecimal string “0x0”.

VI. Generation Process

After the input file is successfully parsed, it is stored in an AST. This tree, especially the decoder part, is transformed in several passes. Each pass either checks the consistency, transforms the tree to change its meaning, or both.

Generation is accomplished using the Visitor design pattern. Transformations are independent visitors, applied to the tree in the specific order. We will describe each transformation very briefly.

A. Instruction Decoder Generation

The instruction decoder node (including its children) of the initial tree constructed by parsing the previously mentioned sample file is shown in the Fig. 1.

The first two passes perform name resolution by creating references between node pairs with the same name.

Then the patterns are joined to form one pattern and one mask per variant. All constant bits are set to one. Bits at the positions of the rules with given length are set to zero.

Edigen supports variable-length instructions (up to 1 kB), but the Java programming language can perform multiway branching only on variables with the maximum size of 32 bits. Thus, the masks and patterns need to be split into smaller pieces (currently, the piece size is set to one byte).

Next, variants are moved from the “top” parts of the tree (assuming the tree is drawn with a root node at the top) to the bottom. This represents the transition from specification to recognition.

Mask or pattern nodes located on the same level and containing the same bit sequences are merged into one node. After this transformation, it is possible to see branching in the tree.

By inspecting the tree, the input file is checked for ambiguity. Generally, we can say that a rule is ambiguous if it is not possible to decide which of its variants matches the given binary input. If an ambiguous rule is detected, Edigen stops the generation process.

An auxiliary pass handle situations when one pattern has multiple child masks. It allows the later generation of “`default`” branches in the code.

Finally, patterns which are children of masks containing only zeros, are removed from the tree because the result of the matching would always be true.

From the final tree (see Fig. 2), the source code of the instruction decoder is generated. Rules are represented by procedure definitions, masks by `switch` statements, patterns by individual `cases` and subrules by method calls.

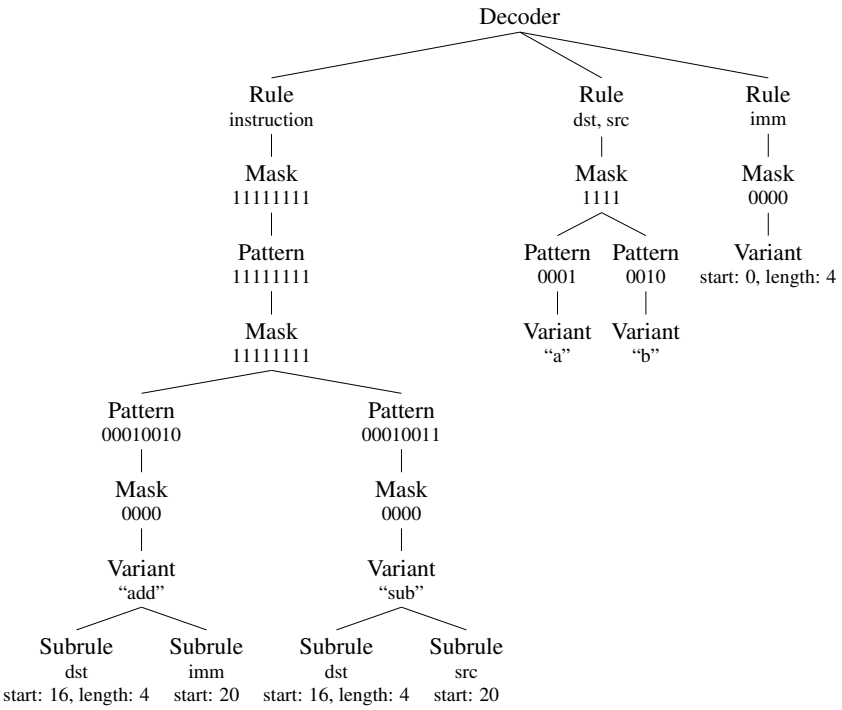


Fig. 2 The final decoder tree, obtained by transformations of the initial AST

B. Disassembler Generation

Generation of the disassembler is much simpler because the majority of disassembler's work is performed at runtime. The disassembler tree node with its children is checked for semantic errors. Then the constants for all formats and values are generated.

C. Templates

The mentioned generation process produces only fragments of code, not compilable on their own. These fragments must be inserted into templates to get the desired functionality. Edigen offers default templates which should fulfill the requirements of almost every user; if not, it is possible to specify custom templates as command-line arguments.

VII. Conclusion

Edigen can largely reduce the amount of time spent by implementing the CPU emulator plugin for emuStudio platform. Only the instruction semantics will be written manually; the instruction decoder and disassembler is produced automatically from the specification.

In the future, it is possible to further optimize the generator without modification of the input specification files.

Acknowledgment

This work was supported by VEGA Grant No.1/0305/11 Co-evolution of the artifacts written in domain-specific languages driven by language evolution.

References

[1] P. Jakubčo, S. Šimoňák, and N. Ádám, “Communication model of emuStudio emulation platform,” *Acta Univ. Sapientiae, Informatica*, vol. 2, 2, pp. 117–134, Nov. 2010.

[2] S. Ducasse, M. Rieger, and S. Demeyer, “A language independent approach for detecting duplicated code,” in *Proceedings of the IEEE International Conference on Software Maintenance*. IEEE, 1999, pp. 109–118.

Distributed Cloud-Centric System Call Tracing Reputation Based (DCSR) Intrusion Prevention System (IPS)

Martin TOMÁŠEK, Marek ČAJKOVSKÝ, Ivan KLIMEK

Department of Computers and Informatics, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Slovak Republic

martin.tomasek@tuke.sk, marek.cajkovsky@tuke.sk, ivan.klimek@tuke.sk

***Abstract* — this paper presents the first theoretic proposal of using system call tracing in a distributed reputation manner. Reputation is a relatively new concept in computer security and as far as the authors are aware of, it was not yet neither used nor proposed to be used in combination with system call tracing. Using such tracing the proposed system is able to observe the behavior of software at run-time thus it does not need to deal with static code analysis and all the methods of code obfuscation. All nodes running the presented system participate in building a reputation database based on what their users run and trust. Information are gathered and distributed via Cloud back-end.**

***Keywords* — attacks, cloud, computer security, intrusion detection and prevention, reputation**

I. INTRODUCTION

The term computer security became a today's buzzword. Almost every day we're hearing about vulnerabilities in software and malicious code that exploits those vulnerabilities. One of the first mentions about computer security was mentioned by J.P. Anderson in early 1980 [1]. A few years later the basic concept of intrusion detection systems was proposed [2]. Intrusion Detection (prevention) Systems – IDS (IDPS) should protect systems/networks and avoid security threats [3][4]. Almost every current IDS is based on one of the 'old' techniques or its combination and slightly differs from the original concept proposed a few decades ago. Unfortunately, practical experience shows that those solutions are not always well working and appear obsolete. Almost all previous security solutions are based on models and techniques that were proposed before rapid Internet and computing power growth. Thus, those solutions hardly reflect the current requirements and available technologies, solutions and knowledge available today. Current challenge is to build autonomous IDS capable of unsupervised learning with low false alarms and discovering zero day attacks.

The system proposed in this paper focuses to be as simple and straightforward as possible. We believe that this novel approach combining the best solutions used so far, low false positives used in rule-based approach and ability to discover zero day attacks used in anomaly-based approach [5-9] can be achieved. We propose to use reputation as a relatively new concept in computer security, combined with cloud computing power and widespread broadband Internet connection available today.

Current trend shows rapid growth of cloud computing becoming widespread phenomena. Services like Gmail, DropBox or YouTube are just tip of the iceberg. Personal computers are no more limited with computing capacities of their own, but rather use the provided resources from the cloud. Many people and companies foretell cloud computing further growing in near future. Cloud computing refers to both the applications delivered as services over the Internet and the hardware and systems software in the data centers that provide those services [10]. Cloud computing users range from individuals and small businesses to Fortune 500 firms and governments. According to a September 2008 survey from the Pew Research Institute, nearly 69 percent of Americans use cloud computing services (such as webmail and online data backup sites) [22]. Also latest boom of private cloud shows that this approach takes the lead. Further studies about cloud computing can be found in [34]. Those features predetermine to use cloud computing in areas with high demand on computing power, availability and resilience which computer security definitely is.

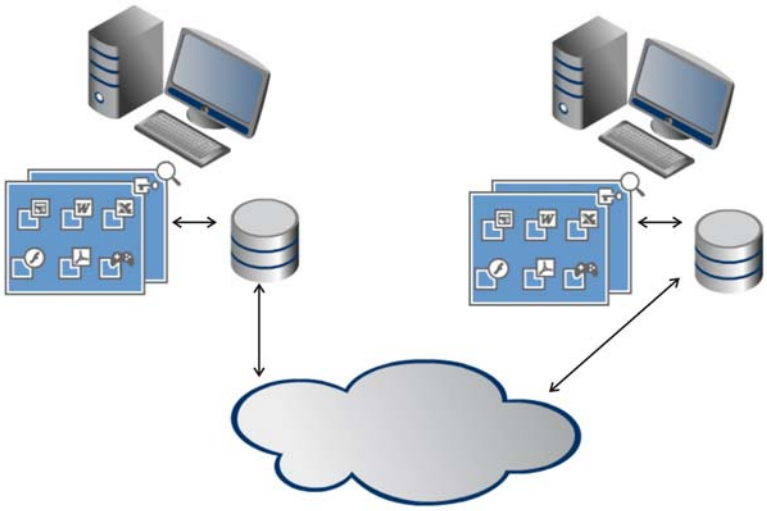


Fig. 1 Simplified architecture of proposed system.

II. INTRODUCING REPUTATION

Considering previously unseen patterns as suspicious and lower its reputation is not a new idea, nearly similar solution SmartScreen was introduced by Microsoft in Internet Explorer (IE) 7 [11]. IE7 protected against URL-based and phishing attacks, IE8 added further URL-based protection [11]. IE9 added another layer of defense, Application Reputation (file reputation) which warns user before running a suspicious application [11-12]. SmartScreen will consider file as suspicious if the Microsoft cloud has not encountered the given file yet, or is known as a dangerous [12]. Microsoft claims that false positive alarms are held to minimum and user warnings won't be needed in 90% of program downloads, thus if the user will see the warning attention should be paid to it [11-13]. An application can also gain positive reputation, for example an application from well-known publisher that has been digitally signed [11]. Those factors lead to implementation of SmartScreen directly into Windows 8 not just in Internet Explorer [14]. Further information about SmartScreen's features and technical details can be found in [15-21].

However this solution works only for files and web pages. Furthermore SmartScreen works with MD5 hashes and digitally certificates, those two authentication methods were exploited in the past - attacks were performed by using stolen valid digital certificates or collision in hashes to generate a binary with the same hash as a valid signature [35][36].

We believe that similar approach deployed on lower level (system calls) of target system could provide better results. We propose a default "deny any" policy. In contrast to a rule-based system which matches known patterns of attacks [7-9]. Our solution considers (same as SmartScreen) patterns of normal behavior and also the pattern of malicious behavior. We are considering files as harmless only those patterns that were observed in the past and were marked as a harmless. This approach makes our system more suspecting toward new, previously undiscovered patterns. After analyzing previously unseen behavior pattern we mark this newly observed pattern as harmless (harmful), and store it to appropriate knowledge database. This approach should rapidly reduce false negatives alarms which behavioral-based IDS suffer from [5-9].

III. KNOWLEDGE DATABASE BOOTSTRAPPING

Following lines outline the first, one of the most difficult phases of our system. This process bootstraps two knowledge databases; GOOD and EVIL. The building consists of following steps:

- Installing known malware samples on a clean virtual machine (available from honey nets etc.).
- Installing binaries that are guaranteed harmless on a clean virtual machine (freeware applications passed latest anti-virus tests etc.).
- Monitoring those test systems – extracting run time system call profiles of the tested binary samples.
- Comparing gathered information against each other.

- This comparison should ensure that the system will be able extract patterns that are typical for malicious and normal behavior and store it in knowledge database.

IV. Binary Analysis - Behavior Monitoring, Extraction and Description

As it was mentioned we believe that the level of observation target system is crucial for this task. Therefore we assume that observation at system calls level is the key. Every user's or program interaction with operating system (packet processing, file accessing, etc.) must be performed via system calls [23-24]. To describe the behavior of a process many techniques were already presented. Those techniques are mainly focused on static or dynamic binary analysis and extracting behavior so called - profiling of a binary. Previous research proves that short sequence of system calls are good discriminator for normal and malicious behavior [25]. But some studies show that exploitation can be done also on this level and evading detection is done by using mimicry attacks [26-27]. Fortunately another study shows how that mimicry attacks can be discovered [28]. By using proper method for capturing program's behavior and techniques described in Knowledge Database Bootstrapping chapter we should be able to describe large amount of normal and abnormal program behavior. We assume that extracted patterns for legitimate (malicious) behavior of one binary will be similar to extracted legitimate (malicious) pattern from another binary cluster, as already shown in [29].

V. System Architecture

Our proposed architecture is based on client server model. Client holds the local database of patterns that is updated after querying cloud. If actual running sequence of binary's system call is missing in client's local database, the cloud is requested. This will save computing power on cloud and network traffic and utilization. Cloud (as it was mentioned) maintains two databases GOOD and EVIL and compares client's requests against those two databases. After the proper match is found, either in GOOD or EVIL database the client is notified with this results. If there is not match neither in GOOD nor in EVIL database client is notified that binary may be dangerous and this binary is send to cloud for further analysis. Cloud stores that information in UNSHURE database. Client's part must use the same techniques to binary behavior description as it was described in Knowledge Database Bootstrapping chapter. Capability of describing binary behavior on client side results from fact that all binaries should run in sandboxed environment, there are several solutions which allows running untrusted software [30-33]. Simplified architecture of proposed system is on Fig.1

VI. Conclusion

In this paper we briefly outlined the problems of current host-based IPS/IDS and presented first theoretical approach for building intrusion prevention system based on relatively new concept, reputation. Our presumptions result from several facts that we described based on prior research and current trends. The most progressive we consider fact that if the reputation concept is well working on URL and application level it is worth considering to deploy this security model on lower level and try to protect whole computer system observing patterns on system calls level. This approach should eliminate the problems related to digital certificate signing used with reputation in SmartScreen nowadays. With computing power and widespread broadband internet available in last years it is no more problematic to run untrusted application in controlled environment, observe their behavior and compare this behavior to previously (un)seen patterns in real time via cloud-back-end.

Acknowledgment

This work was supported by the Slovak Research and Development Agency under the contract No. APVV-0008-10.

References

[1] Anderson - 1980 - Computer Security Threat Monitoring and Surveillance, Fort Washington, PA (Apr. 1980)
[2] Denning - 1987 - An Intrusion-Detection Model
[3] Allison Hrivnak: Host Based Intrusion Detection: An Overview of Tripwire and Intruder Alert, January 29, 200, SANS Institute InfoSec Reading Room
[4] Ning, Jajodia - 2003 - Intrusion detection techniques. The Internet Encyclopedia
[5] Garcia-Teodoro, P. and Diaz-Verdejo, J. and Macia-Fernandez, G. and Vazquez, E. - 2009 - Anomaly-based network intrusion detection: Techniques, systems and challenges
[6] Jamil Farshchi - 12/03/12 - Intrusion Detection FAQ: Statistical based approach to Intrusion Detection - http://www.sans.org/security-resources/idfaq/statistic_ids.php
[7] Andy Cuff - 12/03/12 - Intrusion Detection Terminology (Part One) - http://www.symantec.com/connect/articles/intrusion-detection-terminology-part-one

[8] Andy Cuff - 12/03/12 - Intrusion Detection Terminology (Part Two) - http://www.symantec.com/connect/articles/intrusion-detection-terminology-part-two
[9] S. Jha, M. Hassan - 2001 - Building agents for rule-based intrusion detection system
[10] Michael Armbrust et. al. - april 2010 - A View of Cloud Computing - Doi:10.1145/1721654.1721672
[11] Jeb Haber - SmartScreen® Application Reputation in IE9, 2011/05/17 - [online at] http://blogs.msdn.com/b/ie/archive/2011/05/17/smartscreen-174-application-reputation-in-ie9.aspx
[12] Larry Seltzer - Windows 8: The death of malware? The death of anti-malware?, 2011 - [online at] http://betanews.com/2011/07/20/windows-8-the-death-of-malware-the-death-of-anti-malware/
[13] Larry Seltzer - Will Windows 8 make security software obsolete?, 2011 - [online at] http://betanews.com/2011/09/16/will-windows-8-make-security-software-obsolete/
[14] Steven Sinofsky - Protecting you from malware, 2011/09/15 [online at] http://blogs.msdn.com/b/b8/archive/2011/09/15/protecting-you-from-malware.aspx
[15] Eric Lawrence - IE8 Security Part III: SmartScreen® Filter, 2008/06/02 - [online at] http://blogs.msdn.com/b/ie/archive/2008/07/02/ie8-security-part-iii-smartscreen-filter.aspx
[16] Microsoft, SmartScreen Filter SmartScreen Filter and Resulting Internet Communication in Windows 7 and Windows Server 2008 R2, 2009/12/06 - [online at] http://technet.microsoft.com/en-us/library/ee126149(v=ws.10).aspx
[17] Microsoft, SmartScreen Filter - [online at] http://windows.microsoft.com/is-IS/internet-explorer/products/ie-9/features/smartscreen-filter
[18] SourceDaddy, How the Smart Screen Filter Works - [online at] http://sourcedaddy.com/windows-7/how-smart-screen-filter-works.html
[19] SourceDaddy, SmartScreen - [online at] http://sourcedaddy.com/windows-7/smartscreen.html
[20] John Scarrow - Getting a reputation: How SmartScreen looks at URLs, 2010/03/10 - [online at] http://windowsteamblog.com/windows_live/b/windowslive/archive/2010/03/12/getting-a-reputation-how-smartscreen-looks-at-urls.aspx
[21] Martin Brinkmann - Bypass The Microsoft SmartScreen Filter, 2011/01/28 - [online at] http://www.ghacks.net/2011/01/28/bypass-the-microsoft-smartscreen-filter/
[22] J.B. Horrigan, "Cloud Computing Gains in Currency," 12 Sept. 2008, http://pewresearch.org/pubs/948/cloud-computing-gains-in-currency
[23] JONES, T. M.: Kernel command using Linux system calls, 10 Feb 2010
[24] MITCHELL, M. - OLDHAM, J. - SAMUEL, A.: Advanced Linux Porogramming, ISBN: 0-7357-1043-0, June 2001
[25] Hofmeyr, Forrest, Somayaji - 1998 - Intrusion detection using sequences of system calls
[26] Christopher Kruegel and Engin Kirda - Automating Mimicry Attacks Using Static Binary Analysis
[27] David Wagner and Paolo Soto - Mimicry Attacks on Host-Based Intrusion Detection Systems
[28] Jonathon T. Giffin, Somesh Jha, and Barton P. Miller - Automated Discovery of Mimicry Attacks
[29] Konrad Rieck, Philipp Trinius, Carsten Willems, and Thorsten Holz – 2011 - Automatic Analysis of Malware Behavior using Machine Learning
[30] Bennet Yee, et. al. - Google Inc. - 2009 - Native Client: A Sandbox for Portable, Untrusted x86 Native Code
[31] Cuckoo Sandbox Documentation Book [online at] http://www.cuckoobox.org/doc/0.3.2/latex/CuckooSandbox.pdf
[32] Wei Li, Lap-chung Lam, Tzi-cker Chiueh - Automatic Application-Specific Sandboxing for Win32/X86 Binaries
[33] Mathias Payer, Tobias Hartmann, Thomas R. Gross - Safe Loading-AFoundation for Secure Execution of Untrusted Programs
[34] John Harauz, Lori M. Kaufman, Bruce Potter - JULY/AUGUST 2009 - Data Security in the World of Cloud Computing
[35] Lucian Constantin - 2012/03/12 - Digitally Signed Malware Is Increasingly Prevalent, Researchers Say
[36] Dan Goodin - 2012/06/07 - Crypto breakthrough shows Flame was designed by world-class scientists

Electric Field Influence on Steady Element of Charging Current for High Voltage Insulation Material

Roman CIMBALA, Martin GERMAN-SOBEK, Jozef KIRÁLY

Department of Electric Power Engineering, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Slovak Republic

roman.cimbala@tuke.sk, martin.german-sobek@tuke.sk, jozef.kiraly@tuke.sk

***Abstract* — The paper deals with polarization spectrum of insulation material used in high voltage rotating machines. Low frequency area of polarization spectrum was examined using current response analysis. Isothermal relaxation current analysis is based on current response study of insulation system during charging or discharging process. As a result of this analysis an equivalent Maxwell-Wagner model of dielectrics can be calculated. The influence of voltage on elements of this model in the range of 1V to 1000 V is the aim of the article. An insulation material based on epoxy resin was studies. Comparison of new and old samples was made.**

***Keywords* — insulation system, polarization spectrum, electric field, dielectric spectroscopy, diagnostics of high voltage insulation systems**

I. INTRODUCTION

The high voltage equipment is a very important element in the production process. A breakdown of its insulation system may cause a loss of the production. Therefore the importance of measuring and testing of its electrical properties dealing with its breakdown voltage is increasing with advances of power engineering equipment.

Insulating materials used in high voltage equipment is stressed by temperature, electric field and mechanical vibration mainly. This leads to degradation of its insulating properties, especially electrical strength.

The importance of measurement and testing of dielectric properties of insulating materials is increasing more and more in connection with the new insulating substances that the chemical industry is putting on the market. Within a certain time materials change their physical and chemical specifications.

There are various methods of evaluation and surveillance concerning insulation properties [1]. There is not any characteristic quantity that could determine the state of the insulation system itself now. That is why the set of diagnostic methods is needed. As several methods are applied for diagnostics purposes, the case of several levels of used voltages can be needed.

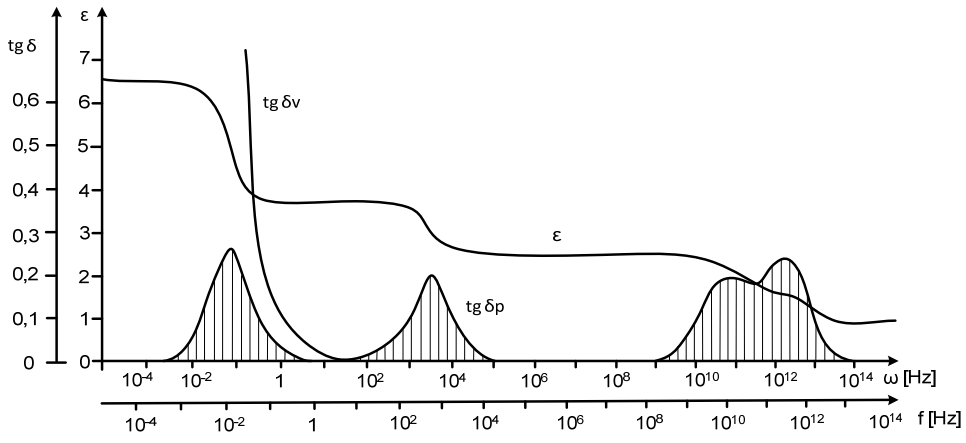


Fig. 1 Polarization spectrum of dielectrics [1]

This paper describes influence of applied external electrics field on insulation material in the range of $2{,}5.10^3$ Vm^{-1} and $2{,}5.10^6$ Vm^{-1}.The application of any diagnostics method based on polarization processes is narrowly connected with bandwidth of observing spectrum. The most information of the common insulation material is in time interval from 10^{-4} to 10^5 s.

Direct methods are based on observing current or voltage time responses. The well-known are polarization indeeces, absorptive or resorptive current analysis, recovery voltage or self-discharge analysis [2], [3] e.t.c.

It was proved that the characteristic shape of the current is changed during the ageing of the insulating material as the electro-physical quality is changed. It is caused by changed of energy spectrum level distribution [4].

Direct measurement of parameters describing changes of energy spectrum of examined material is not possible in the real condition. It is necessary to monitor an external demonstration of dielectric polarization, the absorptive or resorptive current generated applied voltage on the insulating system.

The analysis of the polarization spectrum of the insulating material is based in low frequency domain on the measurements of absorption and resorption currents now in macroscopic level.

The total current consists of three components:

$$i_T(t) = I_0 + i_a(t) + i_g(t) \tag{1}$$

where

$i_T(t)$ – total charging current
$i_g(t)$ – geometrical capacity current,
I_0 – steady current,
$i_a(t)$ – absorption or resorption current.

Decreasing of geometrical capacity current is very fast and it is in 10^{-12} s range. That is why it can be neglected. Then it is possible to consider the total current as sum of elementary relaxation currents exponentially decreasing in the time equation (2), what is fully in accordance with microscopic expression up to the mark distribution power level expressed described equation (1)

$$i_T(t) = \frac{U}{R_0} + \sum_{i=1}^{n} I_{mi} \exp\left(\frac{-t}{\tau_i}\right) \tag{2}$$

$$\tau_i = R_i.C_i \tag{3}$$

$$I_{mi} = \frac{U_0}{R_i} \tag{4}$$

$$I_0 = \frac{U_0}{R_0} \tag{5}$$

where

U – applied voltage,
R_0 – insulation resistance in infinity time ∞,
I_{mi} – amplitude of i-th elementary current,
I_0 – amplitude of steady elementary current,
τi – time constant of i-th elementary current,
n – number of elementary currents.

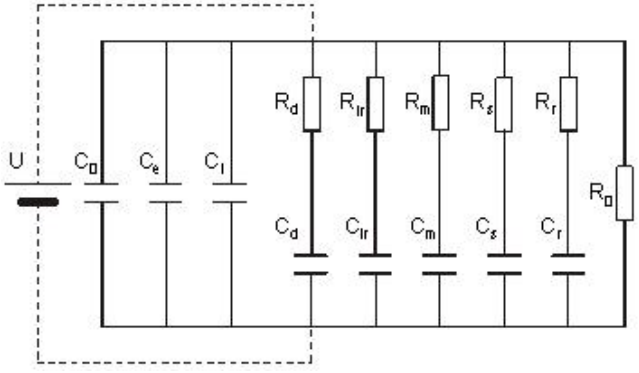


Fig. 2 Maxwell-Wagner's model of dielectrics

Each dielectrics can be replaced by equivalent model consisted of resistance and capacity network. This model, describes system behaviour in time or frequency domain. It has electro-physical base and its physical rendering is possible. The most used network is Maxwell-Wagner model of dielectrics on Fig. 2

Its advantage leads in fact, that it describes relaxation processes in whole dielectric volume. Each element in network characterizes events take place in dielectrics after external electric field influences on it. Over and above, initial condition of dielectrics applying different charges on capacitances can be determined.

In this manner the polarization spectrum of dielectrics can be modelled by scheme, where R_i C_i elements connected in serial mode represented type of polarization with time fixation and can be calculated according to equation (4).

C_0 is geometrical capacity, R_0 is insulation resistivity for dc voltage in infinity time and τ_i is time constant represented elementary polarization event. Capacities and their charges are depended on polarization types. Resistances R represent energy losses of each polarization types.

There is no problem with power supply as in case of alternating diagnostic method. For example, the measurement of capacitive object with capacity about 100 nF and applied voltage 1 kV the requisite current is about 100mA. This is quite huge supply.

II. Steady Element Calculation

Insulation system of electric equipment is tested with several methods. The influence of applied voltage has to be considered especially for measuring and calculation of steady current or insulation resistance of tested object. Resistance is crucial information for insulation system of any electric equipment.

Influence of electric field intensity on Maxwell-Wagner equivalent dielectric model elements was carried out on epoxy-resin samples. The material is wide used for insulation systems of high voltage rotating machine samples of showed frame in Fig. 3 were prepared for testing.

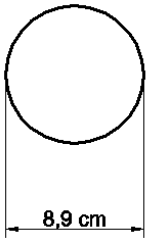


Fig. 3 Prepared samples

The thickness of samples was 4.10^{-5}m. Applied voltage was in the range of 1, 10, 100 and 1000 V. According this the intensity of testing electric field was $2{,}5.10^{3}$ Vm^{-1} , $2{,}5.10^{4}$ Vm^{-1} , $2{,}5.10^{5}$ Vm^{-1} and $2{,}5.10^{6}$ Vm^{-1}. Current in the range of pA were measured. That is why the shielding of samples during measuring was needed. Here is a measuring chamber from permaloy material for electromagnetic screening in Fig. 4.



Fig. 4 Measuring instrument and chamber for electromagnetic shielding of samples

The analysis of total current and its decomposition to elementary currents with amplitudes I_{mi} and time constant of polarization fixation τ_i was carried out using isothermal relaxation current analysis IRC. Utilization of mathematical factorization of total current the steady and time depended elements can be determined. Insulation resistance R_0 can be calculated from steady

current. Number *n* of time depended elements is determined on complexity and material inhomogeneity. Time window of measured current corresponding on time of external electric field application has influence on element number *n* too. Principle of Maxwell-Wagner model elements for polarization spectrum enables the transformation of results reached in time domain to frequency domain [6], [7] or it is possible to use Hamon's approximation as in [8].

As total current was measured using electrometer Keithley 6157B. The instrument was controlled by Agilent Virtual Engineering Environment software through IEEE 488.2 interface. A decomposition process using tools programmed in Matlab was carried out.

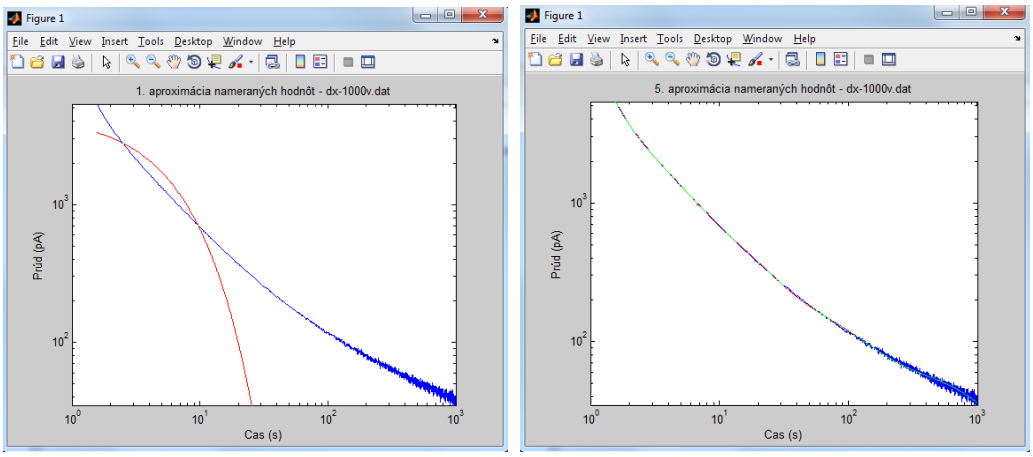


Fig. 5 Decomposition process for 1 and 5 elements

The process is showed in Fig. 5 where decomposition to 1 and 5 elements is calculated. Because of statistic evaluation the analysis for five new and five aged material samples was done. Elementary current magnitudes, time constants of their polarization processes fixations are calculated and from them resistivity and capacity element for them.

The elements from one to seven decomposition processes according to equation (3) were calculated (from I_{10}, I_{11}, τ_{11} and I_{20}, I_{21}, I_{22}, τ_{21}, τ_{22} up to I_{70}, I_{71}, I_{72}, I_{73}, I_{74}, I_{75}, I_{76}, I_{77}, τ_{71}, τ_{72}, τ_{73}, τ_{74}, τ_{75}, τ_{76}, τ_{77}) . The key values were steady elementary current I_{10}, I_{20}, I_{30}, I_{40}, I_{50}, I_{60} and I_{70}.

The influence of applied voltage on steady elementary currents was analysed for new and aged material. The calculated values of currents for new material with voltage as parameter are shown in Fig. 6.

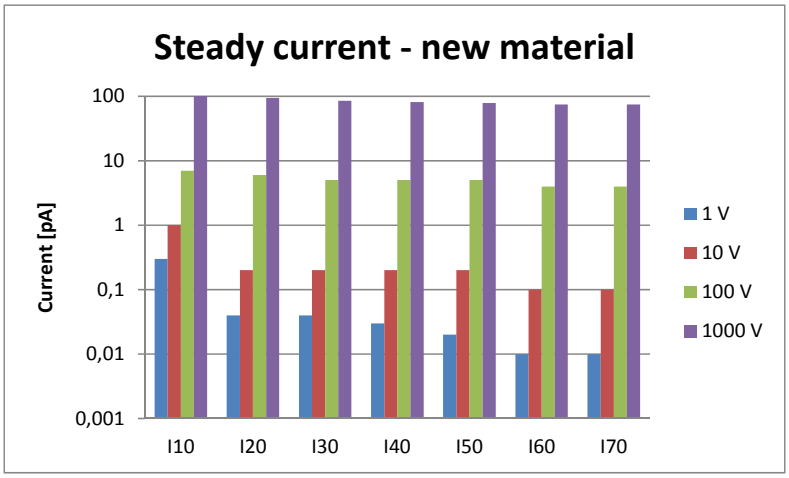


Fig. 6 Steady elementary current for new material with voltage as parameter

The calculated values of currents for aged material with voltage as parameter are shown in Fig.7.

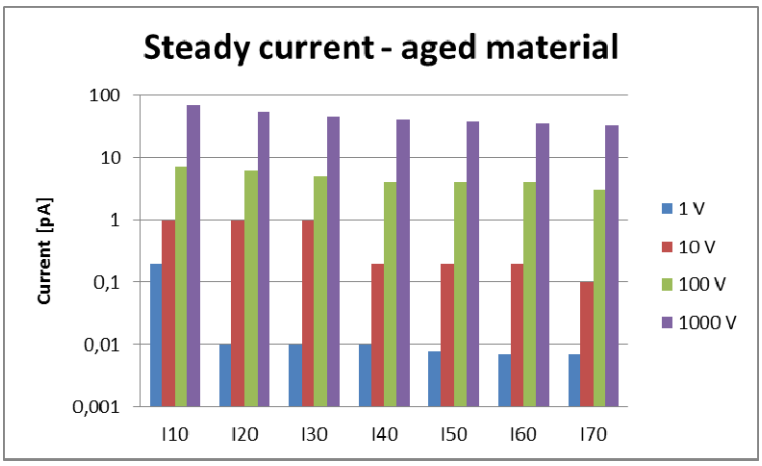


Fig. 7 Steady elementary current for aged material with voltage as parameter

The calculated values of currents for new material depended on applied voltage are shown in Fig.8. The electric field dependence for decomposition to one element according to Maxwell-Wagner equivalent model is not linear, amplitudes grow less than electrical field between 1-100V and then rapidly. Going to decomposition for seven elements the amplitudes raise rapidly form 10 V applied voltage.

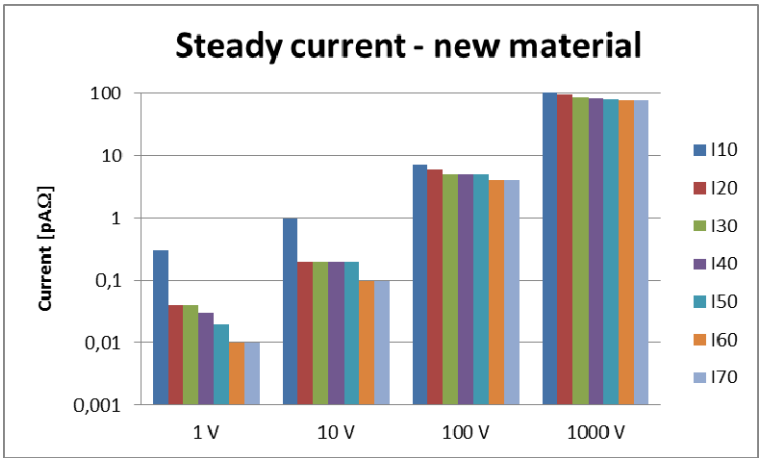


Fig. 8 Steady elementary current for new material for voltage dependence

The calculated values of currents for aged material for voltage dependence are shown in Fig.9.

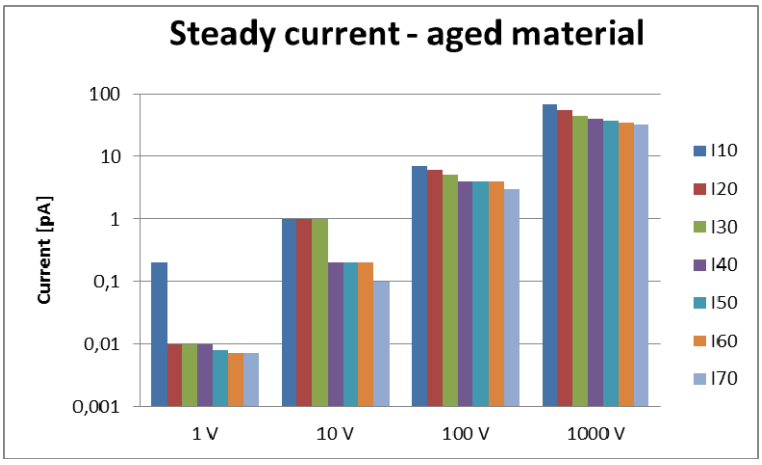


Fig. 9 Steady elementary current for aged material for voltage dependence

The electric field dependence for decomposition to one element according to Maxwell-Wagner equivalent model is still not linear in the beginning, amplitudes grow higher except I_{10}

than electrical field between 1-100V and then they are going to be linear. Going to decomposition for seven elements the amplitudes grow linear or more.

III. CONCLUSION

An influence of electric field on steady elementary currents in Maxwell-Wagner equivalent model was analysed. The influence is not linear and growing of steady current for new material is lower than a growing of an electric field in the range of 10^3 Vm^{-1} up to 10^4 $Vm^{-1}$and higher in the range of 10^4Vm^{-1} up to 10^6 Vm^{-1}. For aged material is situation similar except the intensity of electric field from 10^5Vm^{-1} to 10^6 when it is lower. Aged material has lower current independently of degree of decomposition. It means that according to equation (4) it has higher resistivity. From the first view it is impossible. The fact, that insulation material is complex, has several layers glued together with epoxy resin gives the reason. Epoxy resin is powdering during the ageing process. Delamination process with gap generation appears. Gaps are ideal capacitors with high resistivity and that is why the charging current and its components are decreasing.

ACKNOWLEDGMENT

This work was supported by scientific Grant Agency of the ministry of Education of the Slovak Republic project VEGA No. 1/0487/12. We support research activities in Slovakia / Project is co-financed from EU funds. This paper was developed within the Project "Centrum excelentnosti integrovaného výskumu a využitia progresívnych materiálov a technológií v oblasti automobilovej elektroniky", ITMS 26220120055.

REFERENCES

[1] E. Németh, T. Horvath, "Fundamental of the Stimulation of Dielectric Processes of Insulations", 8-th ISH, Yokohama, Japan 1993
[2] J. Kurimský, B. Dolník: Recovery voltage in ZnO ceramics - 2011. In: Przegląd Elektrotechniczny. Vol. 87, no. 6 (2011), p. 214-217. - ISSN 0033-2097
[3] V., D. Houhanessian, "Measurement and analysis of dielectric response in oil-paper insulation systems", Dissertation work, Swiss Federal Institute of Technology Zürich, Diss. ETH No. 12832, Zürich, 1998
[4] M. Beigert, H.-G. Kranz, D. Kaubisch, D. Meurer, "Computer - Aided Destruction Free Ageing Diagnosis for Medium - Voltage Cables", 8-th ISH Yokohama, Japan, paper 67.11, 1993
[5] T. Leibfried, "Comparability of Dielectric Analysis Results on Power Transformers", Institute of Electric Energy Systems and High-Voltage Technology, University of Karlsruhe Manfred Stach, E.ON Engineering GmbH, Gelsenkirchen, Germany, 2005
[6] T. K. Saha, P. Purkait, F. Muller, "An Attempt to Correlate Time and Frequency Domain Polarisation Measurements for the Insulation Diagnosis of Power Transformer", School of Information technology and Electrical Engineering, University of Queensland, Australia 2005.
[7] W., Katsumi Sakamoto, E. Souza,. D. Das-Gupta, "Electroactive Properties of Flexible Piezoelectric Composites", Materials Research, Vol. 4, No. 3, pp. 201-204, 2001
[8] J. Orságová, P Toman: Cable Thermal Model in Protection System. In In Proceedings of the 9th International Scientific Conference Electric Power Engineering 2008. Brno University of Technology, 2008. s. 197-200. ISBN: 978-80-214-3650- 3.

Modely prevádzok vn elektrizačných sústav a ich porovnanie z hľadiska zavlečenia prepätí do nn rozvodov

Bystrík DOLNÍK, Dominik SMOLKO

Katedra elektroenergetiky, FEI TU v Košiciach, Slovenská republika

bystrik.dolnik@tuke.sk, dominik.smolko@student.tuke.sk

Abstrakt **– Príspevok je zameraný na simuláciu prechodných javov v elektrizačnej sústave vysokého a nízkeho napätia. Porovnávajú sa možnosti zavlečenia prepätia do rozvodov nízkeho napätia z hľadiska typu prevádzky sietí vysokého napätia.**

Kľúčové slová **– EMTP-ATP, prepätie, simulácia, izolácia**

I. Úvod

Technický pokrok, ktorý zastihol aj elektroniku, predstavuje v posledných rokoch kľúčovú oblasť ľudstva. Viac dostupné ceny elektroniky majú dopad na jej veľký rozvoj. Elektroniku treba chrániť pred prepätím, lebo je čoraz viac aplikovaných integrovaných obvodov, ktoré sú citlivejšie na vonkajšie nepriaznivé elektromagnetické vplyvy. Použité elektronické prvky môžu reagovať na prepätia degradáciou. Dôvod na zabezpečenie zariadení pred prepätím je v prvom rade kvôli ochrane života a zdravia, ako aj kvôli ochrane informačných zariadení, kde strata zariadenia by bola z finančného hľadiska malá v porovnaní so stratou cenných informácií.

Zásadným a zložitým problémom sú tranzientné prepätia, ktoré vznikajú ako sekundárna odozva na prechodné javy v prírode ale aj aktivitou ľudí. Účinnými opatreniami pred škodlivými prepätiami sa obmedzí ich vplyv na elektrické a elektronické zariadenia. Ochrana pred prepätím elektronických rozvodov a zariadení je komplexný systém, ktorého súčasťou je bleskozvod, zvody, uzemňovacia sústava, vyrovnanie potenciálov, ekvipotenciálne pospájanie a v neposlednom rade prepäťové ochrany s efektívnym kaskádovým zapojením. V niektorých prípadoch nie je vhodné čakať na vznik prepätí a potom ich obmedzovať, ale je lepšie urobiť opatrenia proti ich vzniku [1].

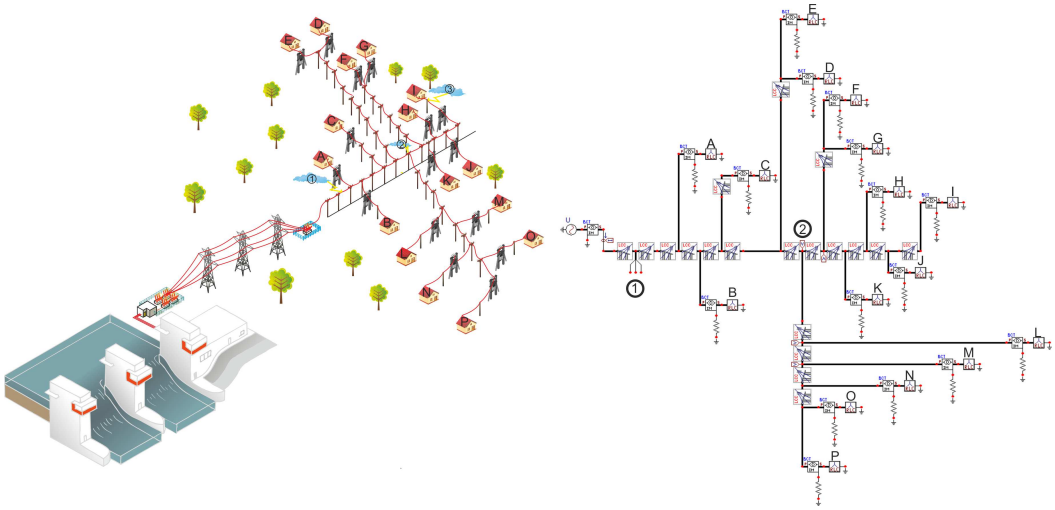
II. Zavlečenie prepätia do rozvodov nn simulované pomocou výpočtovej techniky

A. Schéma a model elektrizačnej sústavy

Simulácia rieši časť elektrizačnej sústavy (ES) na Slovensku. Počas simulácií sa uvažuje s prúdovými impulzmi blesku s rôznymi amplitúdami a s rôznym pôsobením, ktoré má vplyv na veľkosť prepätia v elektrizačnej sústave. Náhradný model obsahuje elektráreň, teda zdroj, z ktorého sa vyvádza výkon cez 110 kV vedenie. Elektrická energia sa pomocou vysokonapäťového transformátora transformuje na napätie 22 kV a potom sa privádza cez distribučné transformátory 22/0,4 kV k spotrebiteľom (do rozvodov nízkeho napätia). Na obr. 1 je znázornená schéma prevádzky 22 kV vn siete s izolovaným uzlom transformátora. Táto schéma sa modifikuje počas simulácií pomocou rozšírenia o jednotlivé prvky ako aj zmenou typu vn siete (zmena uzemnenia uzla transformátora), s cieľom skúmať rozličné stavy na vedeniach. Postupne boli simulované všetky typy prevádzok 22 kV sietí. V nasledujúcej fáze experimentu sa porovnávali jednotlivé typy 22 kV sietí z hladiska zavlečenia prepätia do rozvodov nízkeho napätia [2], [3]. V sieti vn slúžia na zníženie prepätia obmedzovače. V sieti nn sú použité prepätové ochrany (SPD).

V náhradnom modeli na obr. 1, pre lepšie pochopenie obsahu tabuliek a grafov je použité písmenové označovanie uzlov A až P, ktoré poukazujú na jednotlivé záťaže v sieti. Označenie číslicami poukazuje na miesto vzniku prepätí. Body 1 až 3 predstavujú miesta úderu blesku. Body 1 a 2 predstavujú úder blesku do vedenia vn, bod 3 (miesto záťaže I) predstavuje úder blesku do vedenia nn, teda za transformátorom 22/0,4 kV. Úder blesku bol simulovaný impulzom prúdu s amplitúdou 1 kA až 100 kA. Hlavný výboj blesku dosahuje rádovo niekoľko tisíc ampérov. V našich klimatických podmienkach sa hodnoty impulzu prúdu pohybujú v rozsahu 20 kA až 40 kA.

ISBN 978-80-553-0890-6 © 2012 FEI TUKE



Obr. 1 Schéma ES: situačná schéma (vlavo), model simulovaný v programe ATP (vpravo).

B. Nastavenie prvkov elektrizačnej sústavy v programe EMPT ATP

Elektrické zariadenia ES (vedenia, káble, transformátory, odberné miesta, prúd blesku) boli v simulačnom programe EMTP-ATP nastavené podľa reálnych elektrických parametrov.

Nastavenie napájacieho zdroja

Model napájacieho zdroja predstavuje 3-fázový napäťový zdroj. Pri nastavovaní zdroja sa uvažuje so zadávaním maximálneho napätia, teda je potrebné napätie prepočítať z efektívnej hodnoty na maximálnu. Na strane vvn bola nastavená hodnota napätia na

$$U_{\mathrm{m}} = \frac{\sqrt{2} \cdot 110\ \mathrm{kV}}{\sqrt{3}} = 89\,815\ \mathrm{V}\,. \tag{1}$$

Vypočítaná hodnota napätia zdroja sa menila vzhľadom k dovolenej tolerancii napätia v sieti vn a nn, ktorá bola meraná v jednotlivých uzloch elektrizačnej sústavy. Bolo potrebné nastavovať napätie tak, aby hodnoty napätí vn a nn sietí boli tolerancii ± 2 % dovoleného napätia v sieťach. Vypočítane hodnoty napätia na strane vn a nn boli nasledovné:

$$U_{\mathrm{m}} = \frac{\sqrt{2} \cdot 22\ \mathrm{kV}}{\sqrt{3}} = 17\,693\ \mathrm{V} \pm 2\% = (17\,603 \div 18\,323)\ \mathrm{V} \tag{2}$$

$$U_{\mathrm{m}} = \frac{\sqrt{2} \cdot 400\ \mathrm{V}}{\sqrt{3}} = 326,6\ \mathrm{V} \pm 2\% = (320,6 \div 333)\ \mathrm{V}\,. \tag{3}$$

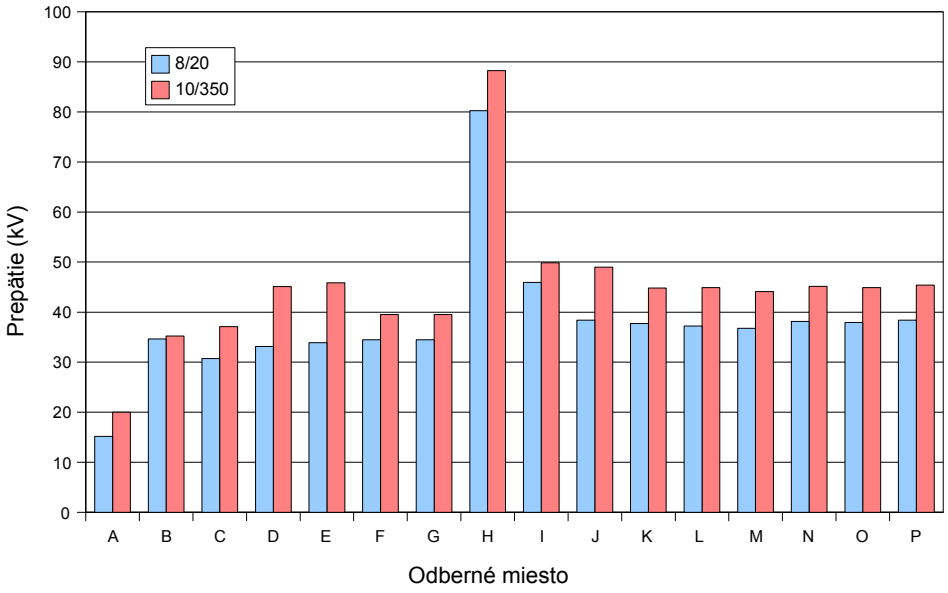
V prípade odchýlky od týchto tolerancií bolo potrebné zvýšiť či znížiť napätie zdroja.

Nastavenie modelu vedenia

Prvok LCC reprezentuje model vonkajšieho vedenia. Pomocou tohto modelu sú namodelované požadované typy vonkajšieho vedenia v simulovanej časti elektrizačnej sústavy. Hodnoty zadávané do modelu sú reálne hodnoty v 22 kV sieti. Uvažuje sa s modelom so sústredenými parametrami, článok π, z dôvodu krátkych úsekov vedení, je frekvenčne nezávislý (platí pre nastavenú hodnotu frekvencie).

Nastavenie modelu transformátora BCTRAN

V simulácií je použitý typ transformátora BCTRAN prostredníctvom ktorého možno simulovať reálne podmienky transformátorov. Do modelu sa zadávajú štítkové údaje transformátora a hysterézna krivka, teda uvažuje sa aj s magnetickým obvodom. Parazitné kapacity a vzájomné kapacity medzi vinutiami nie sú uvažované. Keďže sa s parazitnými a vzájomnými kapacitami medzi vinutiami neuvažuje, treba rozšíriť model transformátora o tieto kapacity s cieľom priblížiť sa reálnym podmienkam v sieti.



Obr. 2 Zavlečenie prepätia do odberných miest pre rozličné tvary prúdových impulzov.

Tabuľka 1
Hodnoty zavlečených prepätí v odberných miestach pre rozličné amplitúdy a tvary prúdových impulzov

Odberné	8/20						10/350			
miesto	1 kA	3 kA	5 kA	10 kA	40 kA	100 kA	1 kA	10 kA	40 kA	100 kA
A	388	636	936	1 686	6 190	15 197	388	2 086	7 970	20 024
B	522	1 211	1 899	3 622	13 958	34 627	500	3 405	13 980	35 226
C	487	1 098	1 708	3 235	12 397	30 720	503	3 830	15 911	37 102
D	471	1 083	1 695	3 226	12 413	33 154	558	4 385	17 934	45 084
E	474	1 098	1 722	3 283	12 642	33 876	565	4 454	18 227	45 837
F	522	1 207	1 894	3 608	13 896	34 472	505	3 874	15 710	39 531
G	522	1 207	1 894	3 608	13 896	34 472	505	3 874	15 710	39 531
H	945	2 542	3 866	8 132	32 089	80 235	1 013	8 943	35 376	88 240
I	615	1 527	2 443	4 733	18 473	45 953	597	4 804	19 827	49 859
J	607	1 488	2 369	4 580	17 862	38 386	590	4 725	19 485	48 993
K	529	1 253	1 979	3 792	14 669	37 729	552	4 350	17 820	44 812
L	509	1 214	1 919	3 681	14 251	37 205	591	4 503	17 939	44 910
M	540	1 257	1 976	3 799	14 783	36 753	547	4 294	16 785	44 101
N	556	1 304	2 053	3 940	15 343	38 153	556	4 388	17 150	45 132
O	553	1 299	2 044	3 919	13 083	37 953	554	4 366	17 073	44 908
P	559	1 316	2 073	3 967	13 764	38 376	559	4 412	17 256	45 409

Nastavenie záťaží ES

Model záťaží predstavuje trojfázový RLC obvod, zapojený do hviezdy. Zadávané hodnoty záťaží bolo potrebné vypočítať. Na strane spotrebiteľa sa uvažuje s indukčnou a odporovou záťažou. Odporové záťaže predstavujú spotrebiče ako žiarovka, ohrievač. Medzi indukčné patria asynchrónny motor (AM), a spotrebiče s transformátormi.

III. Výsledky a diskusia

V simuláciách sú porovnané prúdové vlny nielen z pohľadu ich maximálnych hodnôt prúdov ale aj z pohľadu ich časového pôsobenia. Simulácia pomocou prúdovej vlny 10/350 reprezentuje priamy uder blesku. Prúdová vlna s tvarom 8/20 reprezentuje nepriame účinky blesku. Graf na obr. 2 poukazuje na to, aký vplyv má účinok blesku na vznik prepätia a to nielen z pohľadu amplitúdy ale aj z pohľadu jeho dĺžky pôsobenia. Porovnané sú dva základné tvary impulzov prúdu: 8/20 a 10/350. V tabuľke 1 sú uvedené hodnoty prepätí pre rozličné amplitúdy a tvary prúdových impulzov simulovanej ES zavlečené do bodu 1, pozri obr. 1.

V tabuľke 2 sú uvedené hodnoty zavlečených prepätí v 22 kV sieti, merané v jednotlivých uzloch (odberných miestach). Porovnané sú navzájom zavlečené prepätia podľa typu prevádzky vn siete. Kompenzovaná sieť reprezentuje prevádzku vn siete cez Petersonovú tlmivku. Šírenie prepätí vo vn kompenzovanej sieti a v sieti izolovanej je porovnateľné. Naopak v sieti vn s uzemneným uzlom transformátora cez odporník sú rozdiely väčšie. Pomocou simulačného programu EMTP-ATP bolo zistené, že typ prevádzky siete vn nemá vplyv na veľkosť zavlečenia prepätí do nn siete.

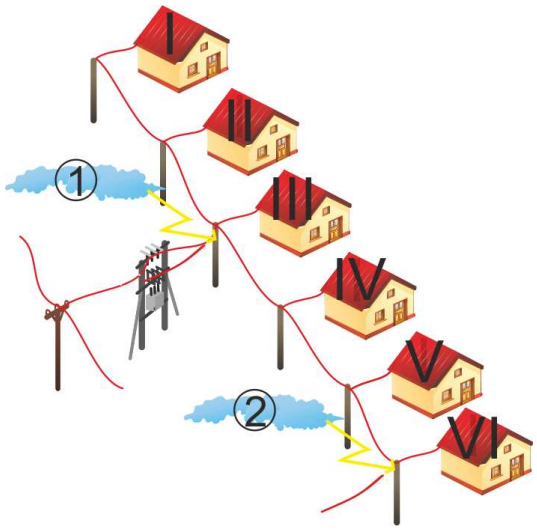
Tabuľka 2
Hodnoty zavlečených prepätí do jednotlivých odberných miest vn siete pri údre blesku bez uvažovania zvodičov prepätia

Typ vn siete	Prepätie v označených miestach (kV)[a]							
	1	1A	A	ED	2	FG	L	M
Izolovaná	23,34	20,39	20,33	13,61	12,27	15,76	13,08	15,61
Uzemnená cez odporník	19,99	16,42	16,48	10,64	9,81	12,21	10,95	12,64
Kompenzovaná	23,34	20,41	20,34	13,58	12,21	15,71	13,02	51,51

[a]Použitý impulz prúdu s tvarom 8/20.

V praxi sa inštalujú zvodiče prepätí s cieľom znížiť účinky prepätí na elektrické zariadenia, čím je zavlečenie prepätí oveľa menšie, pozri hodnoty v tabuľke 2. Inštalácia obmedzovačov prepätia je potrebná všade tam, kde dochádza k zmene vlnovej impedancie.

Na obr. 3 je zobrazený detail záťaže v odbernom mieste E na nn strane ES. Pozostáva z viacerých menších záťaží (domov). Na obrázku sú označené rímskymi číslicami I až VI. Navzájom sú spojené paralelne galvanicky pomocou 50 m dlhých vedení. Celková záťaž v odbernom mieste E je prepočítaná na šesť menších záťaži s rovnakou hodnotou.



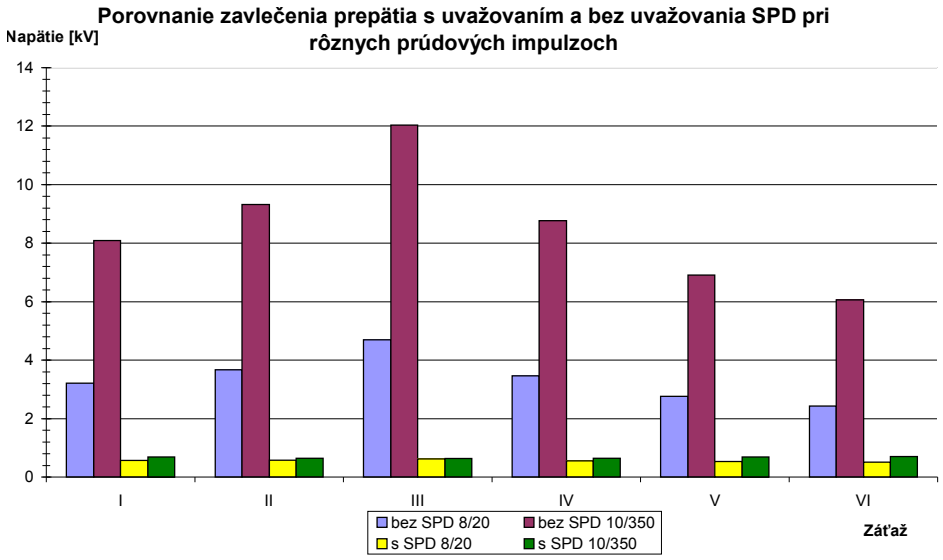
Obr. 3 Detail modelu odberného miesta E.

V experimente boli simulované rozličné prípady úderu blesku do rôznych odberných miest bez inštalovanych ako aj s inštalovanými prepäťovými ochranami. V prípade prúdovej vlny s amplitúdou 1 kA a tvarom 8/20 je zavlečené prepätie menšie oproti prúdovej vlne s tvarom 10/350 s rovnakou amplitúdou. Z toho vyplýva, že priamy úder blesku má podstatne nepriaznivejšie účinky ako nepriame účinky blesku. Pri nepriamych účinkoch blesku, v prípade inštalovania SPD sa šírené prepätia po vedení postupne tlmia. V prípade priamych účinkov boli pozorované sekundárne javy spôsobené výbojom blesku, ktorého energia nebola ešte absorbovaná prepäťovými ochranami.

Na obr. 4 je uvedený graf, v ktorom sú porovnané vypočítané prepätia na záťaži III s použitými a bez použitých SPD; ďalší parameter bol tvar impulzu prúdu. Na grafe vidno účinok SPD v znížení prepatí, čím sa chránia inštalácie a elektrické zariadenia v mieste odberu elektrickej energie. V prípade uvedenom na obr. 4 bol inštalovaný jeden ZnO zvodič prepätia a to v mieste záťaže III. Z grafu vidno (porovnaním amplitúd prepätí), že v tomto prípade by boli chránené aj zvyšné domy pred veľkými škodami na majetku, pretože amplitúda prepätí na nn strane prevyšuje maximálne napätie v nn sieti len približne o 100 V.

IV. Záver

V elektrizačnej sústave má najväčší vplyv na namáhanie izolácie a na degradáciu zariadení atmosférické prepätie. Predísť škodám možno iba pomocou komplexného systému, ktorý eliminuje nežiaduce účinky prepätí. Simuláciou časti ES (vn a nn siete) pomocou programu EMTP-ATP bolo skúmané zavlečenie prepätia do rozvodov nízkeho napätia. Vzájomné porovnanie prepätí vo vn a nn sieti je založené na komparačnej metóde spočívajúcej v stanovení amplitúdy prepätia s inštalovanými a bez inštalovaných zvodičov prepätia a ich následného zavlečenia do rozvodov nízkeho napätia.



Obr. 4 Porovnanie zavlečenia prepätia s uvažovaním a bez uvažovania SPD pre rôzne impulzy prúdu blesku.

Výsledky simulácií potvrdzujú dôležitosť inštalovania zvodičov prepätí na vn strane. Bez ich inštalácie by dochádzalo k značným škodám na majetku a na elektrických zariadeniach. Úroveň zavlečenia prepätí je porovnávaná z pohľadu zapojenia uzla transformátora vo vn sieti. Zavlečenie prepätia na vn vedeniach bez inštalovaných obmedzovačov prepatia závisí od typu prevádzky uzla transformátora. Na zavlečenie prepätia do rozvodov nízkeho napätia nemá spôsob prevádzky uzla transformátora vplyv. Aj keď distribučné transformátory zvedú určitú časť prepätia, škody na distribučných transformátoroch bez inštalovania obmedzovačov by boli veľmi veľké.

Poukázané je na nutnosť inštalovania SPD v rozvodoch nízkeho napätia blízko jednotlivých spotrebiteľov, pretože blesk môže udrieť aj do rozvodov nízkeho napätia a v takom prípade už zvodič prepätia inštalovaný v sieti vn nepomôže chrániť izoláciu nn elektrických zariadení. Pri simuláciách úderu blesku je tiež porovnávaný vplyv priamych a nepriamych účinkov blesku s rovnakou amplitúdou.

Poďakovanie

Autori vyjadrujú poďakovanie za podporenie tejto práce Vedeckej grantovej agentúre Ministerstva školstva Slovenskej republiky a Slovenskej akadémie vied v rámci projektu VEGA-1/0487/12.

Literatúra

[1] J. Burant, *Blesk a přepětí: Systémová řešení ochran*. FCC Public, 2006.

[2] M. Costea and B. Nicoara, "The effects of lightning induced overvoltages on low voltage power networks," in *PowerTech, 2009 IEEE Bucharest*. IEEE, Oct. 2009, pp. 1–6. [Online]. Available: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5281814

[3] *ABB Surge Arresters - Produkty VVN*, ABB HIGH VOLTAGE TECHNOLOGIES LTD, Division Surge Arresters., Wettingen, Switzerland: ABB, Aug. 2008. [Online]. Available: http://www.abb.com/product/cz/9AAC710009.aspx?country=00

IRC Analysis of Degraded XLPE Cable

Martin GERMAN-SOBEK, Jozef KIRÁLY, Roman CIMBALA

Department of Electric Power Engineering, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Slovak Republic

martin.german-sobek@tuke.sk, jozef.kiraly@tuke.sk, roman.cimbala@tuke.sk

***Abstract* — The change of dielectric parameters points to the aging or defect of HV equipment insulation. Cross-linked polyethylene (XLPE) is the world's preferred modern insulation for HV cables. This article deals with the measurement of current-time curve of degraded specimen of XLPE cable. In the article are also describes the processes and reasons of ageing XLPE cables. The measurement was carried out using isothermal relaxation current analysis (IRC analysis) with using a Keithley 617 electrometer. IRC analysis represents a non-destructive one-way method suitable for measuring the dielectric parameters of polymeric composites during aging.**

***Keywords* — IRC analysis, XLPE cable, destruction free diagnostic, ageing of XLPE, cable insulation**

I. INTRODUCTION

Insulation system is an important and sensitive part of any electrical equipment. In the present is dominated increased demand for electricity and therefore is required greater security and reliability for high-voltage electrical equipment. These requirements relate mainly to the insulation system of equipments because damage of insulation can lead to equipment failure and other disorders.

Cross-linked polyethylene (XLPE) is widely used as electrical insulation material for high voltage power cables because of its excellent electrical, physical and chemical properties. These cables are usually exposed to several kA current and hundreds of kV voltage and they are critical parts of the transmission infrastructure. Therefore, is expected their high resistance against possible failures during their lifetime [10].

The cables are permanently exposed to thermal aging during operation. It may cause irreversible damage of cable insulation. The primary initiators for the degradation of cable insulation are high currents, voltage, mechanical, chemical and thermal stress and pollution of environment. Interaction various factors together may significantly speed up the degradation processes. Process of aging of insulation is the most acting on the quality of insulation. Therefore, it's necessary to assess degradation and insulating state of cables.

One of methods for detection of insulating state of XLPE insulation is IRC analysis. This method is a non-destructive method. This paper presents results of the analysis of XLPE insulated power cables at an advanced stage of ageing.

II. METHOD OF ANALYSIS

IRC– analysis is one of the DC non-destructive methods. This method offers possibility to investigate the degradation processes in the material. The state of insulation can be determined by measuring the relaxation currents in the time domain.

Real dielectrics contain electric charges bound by electrostatic forces and there are also free charge carriers, which create undesirable conductivity. Dielectric polarization is a phenomenon that occurs under the effect of electric field and causes the movement of free charge carriers [2], [7].

If dielectric material is inserted in the electric field of intensity $E(t)$, this material will be polarized. Electric polarization $P(t)$ in dielectric materials can be divided into two parts, namely the fast and slow polarization [3].

Slow polarization processes are measured by charging and discharging currents.

Current $i(t)$, which flows through the dielectric when exposed to electric field of intensity $E(t)$ can be expressed as:

$$i(t) = C_0\left[\frac{\sigma_0}{\varepsilon_0}u(t) + \varepsilon_\infty \frac{du(t)}{dt} + \frac{d}{dt}\int_0^t f(t-\tau)u(\tau)d\tau\right] \quad (1)$$

where:
C_0 - geometric capacity of the measured object

Polarizing current $i_{pol}(t)$ can be expressed as:

$$i_{pol}(t) = C_0 U_C\left[\frac{\sigma_0}{\varepsilon_0} + f(t)\right] \quad (2)$$

As a result of a DC voltage U_C, polarizing current is composed of two parts. The first part is related to conductivity of the investigated object and the second part is related to activation of different polarization processes inside the investigated material [3], [7].

If the DC voltage UC is removed at the time T_C and the material is grounded, depolarization current $i_{depol}(t)$ starts to flow through the material.

$$i_{depol}(t) = -C_0 U_C\left[f(t) - f(t+T_C)\right] \quad (3)$$

where:
$f(t)$ - monotonically decreasing function [3]

A course of current response after connecting and disconnecting the DC voltage (charging and discharging current) is shown in Figure 1.

The total polarizing current flowing through the dielectric can be following assumption of existence the independent Debye's polarization processes and Maxwell-Wagner equivalent model, expressed as the sum of currents with exponentially decreasing amplitude with definite-time component, which represents the individual of the polarization occurring in the material by the equation:

$$i(t) = \frac{U}{R_0} + \sum_{i=1}^{n} I_{mi} \exp\left(\frac{-t}{\tau_i}\right) \quad (4)$$

where:
U – applied DC voltage
R_0 – DC insulation resistance after an infinitely long time
I_{mi} – amplitude of i-th component of Debye's elementary current
τ_i – relaxation time constant of the i-th component [7]

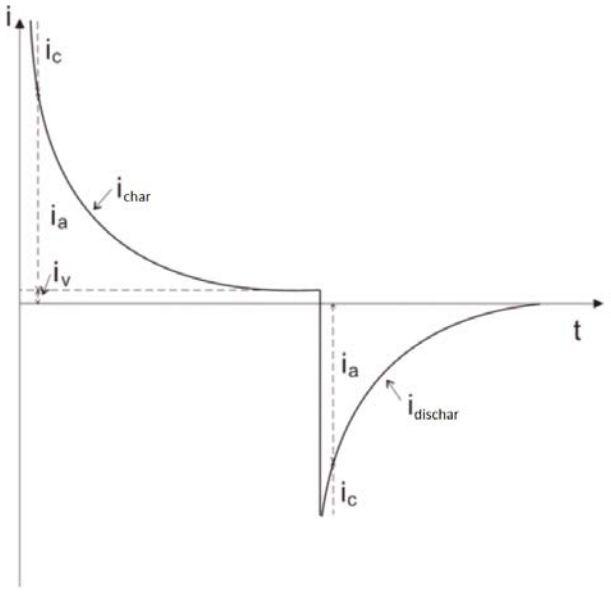


Fig. 1 Current response after connecting and disconnecting the DC voltage [4]

The parameters I_{mi}, τ_i are correlated with the material properties. The time constant τ_3 is related to water degradation of the cable insulation [11].

Number of members of expansion n is dependent on the structure and homogeneity of the dielectric and the time at which the polarizing action are measured [6], [7].

The time characteristic of the isothermal relaxation current varies significantly with the residual strength of polymeric composite specimens. Performing the IRC-Analysis every relaxation current component can be associated with a time function [9].

During aging, it is observed increase of slower polarization process at using IRC analysis. This increase is captured to measurement technique through a separate determination of current-relaxation amplitudes. The ageing leads to the amplified formation of the current amplitudes in a time period greater than 100 seconds [7].

III. Processes of Ageing in XLPE

Polyethylene in its cross-linked forms is the insulation material of a XLPE and its a single homogeneous dielectric.

XLPE cable ageing has been studied for nearly 40 years and many methods have been proposed to evaluate the properties of XLPE [13]. Ageing in a polymer changes the electrical, physical, mechanical and morphological properties of the insulation. All these properties are influencing the dielectric parameters and characteristics of insulation [9].

Ageing of XLPE cables is related to the temperature of the insulation. All XLPE cables contain antioxidants which protect the XLPE from oxidation during the extrusion and cross-linking process, and also during the service life of the cable. The rate at which the antioxidant is used up is dependent on temperature. The normal maximum operating temperature of XLPE cables is 90°C. Increasing the operating temperature of the cables will increase the rate at which the anti-oxidant is used up and hence reduce the service life. Small increase in temperature has a significant impact on the ageing of the XLPE. The cable will be subject to stress cracking and electrical failure at positions of mechanical stress [12].

Tests have shown that XLPE cables can operate at a temperature of 105°C for a limited time without significantly reducing the service life of cables. As the temperature increases XLPE softens and the mechanical performance of the XLPE reduces considerably. The maximum overload temperature of XLPE is limited to 105°C [7], [12].

When the cables are ageing in hydrothermal condition for a period of time, the mechanical and electrical properties of the XLPE insulation may degrade, which may lead to an apparent deterioration in dielectric performances such as electric conductivity, permittivity and dielectric loss. The physical form and chemical structure of XLPE cables will be altered. During thermal ageing, several structural changes occur such as variation in crystalline, chain scission and variation in heat of fusion and in melting point [13].

Quantitative assessment of aging or damage of the XLPE insulations can be separated assessment of the relaxation mechanisms.

IV. Experiment

A. Measurement

The measurement was performed by the IRC analysis using Keithley 617 electrometer for one specimen of HV cable. Circuit configuration for measurement is shown in Figure 2, where C_X is measured specimen of cable.

The cable has been a power cable with an aluminium core and XLPE insulation. The status of cable insulation has been operationally aged. Specimen of cable was approximately 0.25 m long and protective cable jacket and semiconducting layer have been removed (5 cm at both ends) and shielding taken out. Time of measurement was 1000 s. During this time, it is possible to measure by IRC analysis only 7 types of polarization.

Specimen of cable was connected to test voltage (100V) by Keithley 617 electrometer and the specimen was subjected to affect of this electrical field for 1000 seconds. After the first measurement of data, the specimen was thermally degraded at 90°C for 24 hours. Then measurement was repeated. By means of software were obtained the currents I_{mi} and the times τ_i.

Aim of measurement has been comparison and evaluation of measured dependence of charging current and obtained times τ_i of operationally aged specimen of cable before and after the additional degradation.

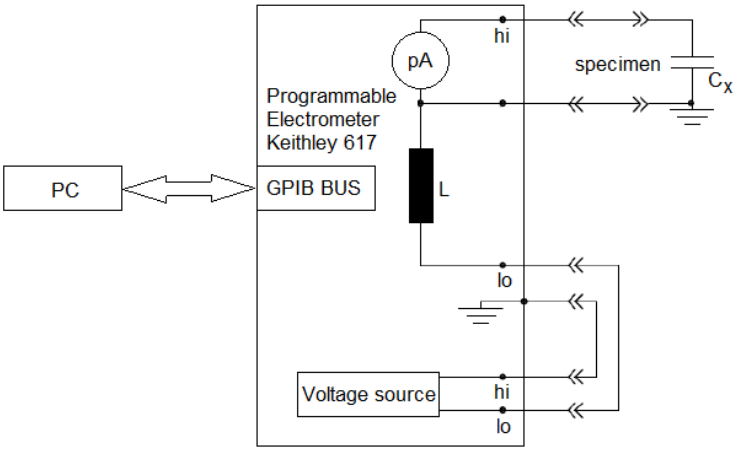


Fig. 2 Schematic diagram of measuring circuit

B. Evaluation of measurement

Figure 3 shows the measured time dependences of charging currents of cable specimen. The displayed waveforms have visible difference.

As shown the Figure 3, the specimen of degraded cable reached larger values of current approximately in time to 2 second. From this point, the specimen of degraded cable reached lower values of current as from 24 hour degradation. In time about 2 seconds in both cases of measurement was recorded the deflection of the downward trend of charging current (moderate increase followed by a steep decline and again a slight increase of current). This deflection was probably caused some polarization phenomena taking place in that time.

From measured dependence of charging currents were calculated constants of elementary currents and relaxation times for seven polarization processes. Calculated data are shown in Table 1.

For further interpretation was excluded the development of these amplitude of current and relaxation time, which included values less than or equal to 0 or greater than 10^7, because it is not possible that currents were negative values.

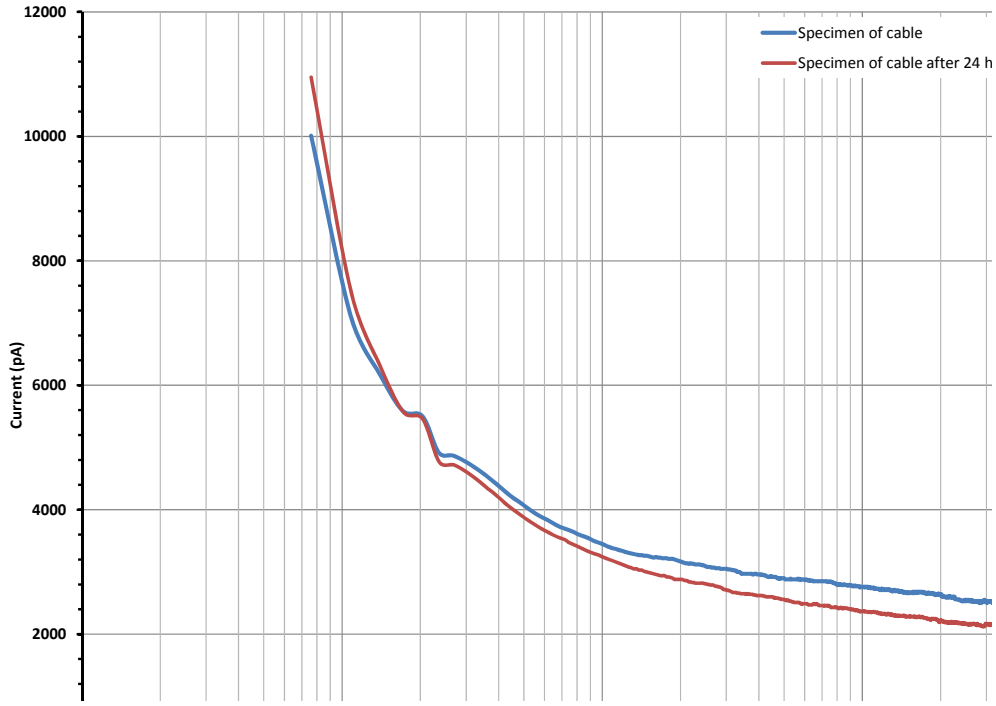


Fig. 3 I-t characteristics of cable specimen

As shown the Table 1, values of time and current for the 7th approximation were incorrect, since they reach negative values. For this reason, the values τ_i and I_{mi} for n=6 polarization processes were used in analyzing, thus the 6th approximation.

For analyze the quality of materials is an important analysis of stabilization time constants of elementary currents. Comparing the time constants obtained from Table 1 shows, that there has been a slight increase in values of times τ_{11}, τ_{22}, τ_{33} and τ_{44} after 24 hour degradation at 90°C. From this it can be concluded that 24 hour degradation caused an increase in these values. On the other hand there was a decrease in values of times τ_{55} and τ_{66}.

Table 1
Constants of elementary currents and relaxation times for seven polarizing processes

	Specimen of cable		Specimen of cable after 24 hours degradation	
Approximation	Relaxation time τ_i (s)	Relaxation current I_{mi} (pA)	Relaxation time τ_i (s)	Relaxation current I_{mi} (pA)
1	3611.25	2859	3827.35	2473
2	3.8	5538	4.23	6034
	4339.33	2776	5120.6	2371
3	1.67	8248	1.12	12719
	76.43	791	42.38	1238
	8049.96	2582	10256.83	2221
4	0.37	34056	0.4	36043
	4.83	2964	5.14	2946
	143.36	644	91.47	763
	13325.74	2481	20988.64	2141
5	0.27	61135	0.34	50405
	2.81	3575	4.72	3157
	19.47	673	59.28	606
	259.66	648	212.79	298
	153831.56	2318	60493.43	2078
6	0.27	62155	0.31	55740
	2.82	3583	2.68	3608
	19.71	661	17.19	921
	246.52	591	139.7	555
	568.91	82	498.84	66
	1037303.25	2294	805466.04	2047

V. CONCLUSION

Aim of this paper was to comparison and evaluation of measured dependence of charging current and obtained times τ_i of operationally aged specimen of XLPE cable before and after the additional degradation. It is known that ageing changes the properties of insulation. The quality of insulation and her electrical parameters are dependent on a number of factors that affect the insulation during operation. For measurement was chosen method of IRC analysis, which offers possibility to investigate the degradation processes in the material.

It was found a change in *I-t* characteristic for specimen of XLPE cable before and after 24 hour degradation at 90°C. It was also found change in calculated elementary currents and time constants.

Comparison of time constants τ_{11}, τ_{22}, τ_{33} and τ_{44} shows increased values for specimen after additional degradation. This confirms that this specimen is more degraded and aged. It is possible to confirm that degradation led to changes in the polarization processes occurring in this specimen. The measured *I-t* characteristics show a change in values of current for specimen before and after degradation. It is possible thus to assume that the trend rate of polarization processes occurring in insulation of specimen was different.

In conclusion, the measurement confirmed the suitability of IRC analysis for investigate the degradation of cable insulation system. It is not possible exclude the surrounding of a disturbance during the measurement, which could affect the measured results. Measured specimen is suitable for next investigation.

ACKNOWLEDGMENT

We support research activities in Slovakia / Project is cofinanced from EU funds. This paper was developed within the Project "Centrum excelentnosti integrovaného výskumu a využitia progresívnych materiálov a technológií v oblasti automobilovej elektroniky", ITMS 26220120055.

This work was supported by scientific Grant Agency of the ministry of Education of the Slovak Republic project VEGA No. 1/0487/12.





REFERENCES

[1] K. Marton, Elektrotechnické materially (Fyzika dielektrík), ESVŠT, Košice, 1971.
[2] K. Marton, et al.: Situation analysis of electric field on boundary-line of solid state electrode and gas. In: In: Przegląd elektrotechniczny. Vol. 86, no. 8 (2010), pp. 307-311. ISSN 0033-2097.
[3] W. S. Zaengl, "Dielectric Spectroscopy in Time Domain for HPV Power Equipment, Part I: Theoretical Consideration," IEE Transaction on Dielectrics Electrical Insulation, Vol. 19, No. 5, 2003, pp. 5-19.
[4] Model 617 Programmable Electromer, Instruction Manual, Keithley Instruments, Inc., Cleveland, Ohio, USA, 1984
[5] T. Palai-Dany, "Dielectric Relaxation Spectroscopy with Keithley 617 Electrometer, EEICT, CVUT, Brno, 2006.
[6] R. Cimbala, "Dielektrické vlastnosti XLPE káblov" In: Starnutie Elektroizolačných systémov. Vol. 1, No. 1, 2006, pp. 14-16.
[7] R Cimbala, Starnutie vysokonapäťových izolačných systémov, TUKE, Košice, 2007, 188p.
[8] M. Muhr, R. Schwarz, S. Jaufer, "Electrical Measurements as Diagnostic Tool for HV-Insulations," Graz, 2005, pp. 1-11.
[9] G. Hoff, H. G. Kranz, "Interpretation of Dielectric Response Measurement Data from Service Aged XLPE-Cables," IEEE 7th International Conference on Solid Dielectrics, Eindhoven, June 25-29, 2001, pp. 381-384.
[10] J. C. Fothergrill et al., "The Measurement of Very Low Conductivity and Dielectric Loss in XLPE Cables," IEE Transaction on Dielectrics Electrical Insulation, Vol. 15, No. 5, 2011, pp. 1544-1553.
[11] Oyegoke, B.S. et al.: Condition Assessment of XLPE Insulated Cables Using Isothermal Relaxation Current Technique. In: Power Engineering Society General Meeting, IEEE, 2006, pp. 6. ISBN: 1-4244-0493-2.
[12] Ageing of XLPE Compounds, Brochure, General Cable Australia Pty Ltd.
[13] Jianying, L. et al.: The Effect of Accelerated Water Tree Ageing on the Properties of XLPE Cable Insulation. In: IEEE Transaction on Dielectrics and Electrical Insulation, Vol. 18, 2011, pp. 1562-1569.

Konštrukcia modelu veterných elektrární

Slavomír HERBERT, Ján TKÁČ

Department of Electric Power Engineering, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Slovak Republic

slavo.herbert@gmail.com, jan.tkac@tuke.sk

***Abstrakt* — Táto práca sa zaoberá využívaním veternej energie na výrobu elektrickej energie, pomocou veterných elektrární s vertikálnou osou. Inštalované výkony veterných elektrární narastajú s postupným ubúdaním fosílnych palív. V súčasnosti sú najviac používané veterné elektrárne, ktorých turbíny majú horizontálnu os otáčania. Cieľom práce bolo skonštruovať plne funkčné modely veterných turbín s vertikálnou osou. Vlastnosti jednotlivých modelov veterných turbín boli overované laboratórnymi meraniami a graficky vyhodnotené. Výsledky poukazujú, že konštrukcia rotorov má vplyv na výsledné výstupné parametre. Najviac sa osvedčili viacstupňové rotory.**

***Kľúčové slová* — veterná energia, vertikálny rotor, meracie zariadenie, konštrukcia modelu**

I. Úvod

V súčasnosti sa ľudstvo zameriava na využívanie obnoviteľných zdrojov energie, ktoré neznečisťujú životné prostredie. Takýmito zdrojmi energie sú slnko, vietor, voda, ktoré sa nachádzajú všade okolo nás. Obnoviteľné zdroje energie a výroba elektrickej energie pomocou nich je výhodná najmä kvôli tomu, že nenarušujú termodynamickú strabilitu planéty a znižujú produkciu emisií [1]. Ich výhodou je možnosť výroby elektriny distribuovaným spôsobom na ktoromkoľvek mieste na planéte bez spolupráce s elektrickou sieťou, alebo pracujúce do autonómnej siete [2]. Využiteľnosť vetra siaha do dávnej minulosti, kedy ho ľudia začali využívať na mletie obilia za pomoci veterných mlynov, využíval sa aj v námorníctve na pohon lodí a pod. V dnešnej dobe je vietor využívaný najmä na výrobu elektrickej energie a veterné elektrárne majú inštalované výkony až na úrovni MW. Ich výhodou je, že vyrábajú elektrickú energiu získanú zo zdroja , ktorý je zdarma [3]. Veľmi dobre sa dopĺňajú s fotovoltickými zdrojmi čím sa v spolupráci s akumuláciou vyrobemej elektriny dosahuje viacvalentnosť systému [4], [5], [6].

Náplňou práce bolo navrhnúť a skonštruovať malé veterné turbínky rôznej konštrukcie a laboratórnymi meraniami zistiť ich vlastnosti a porovnať ich. Modely veterných turbín rôznych tvarov budú v konečnom dôsledku tvoriť laboratórnu veternú farmu, ktorá bude slúžiť aj ako ukážka prevádzky veterných elektrární.

II. Konštrukcia Malých Veterných Turbín

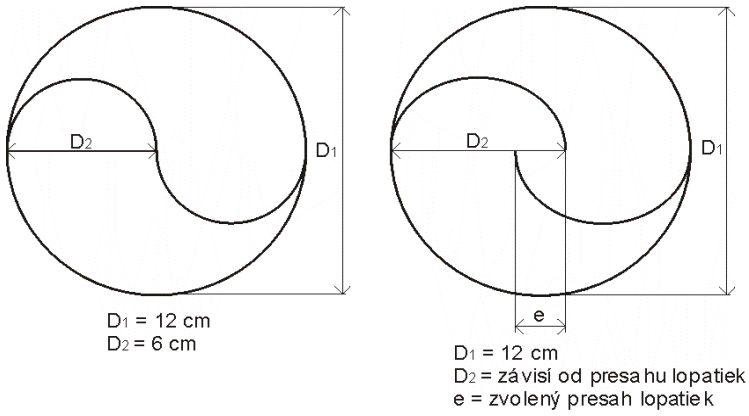
Návrh modelov veterných turbín vychádzal z rotorov s vertikálnou osou otáčania, ktoré sa vyznačujú tým, že majú dobrý záberový moment a nie je ich potrebné natáčať do smeru vetra. Rozbiehajú sa pri nízkych rýchlostiach vetra a dobre spracúvajú aj turbulentný a nárazový vietor. Konštrukčne bolo potrebné zhotoviť modely vertikálnych veterných turbín rôznych druhov, ktoré budú tvoriť laboratórnu veternú farmu. Pri konštrukcii modelov veterných turbín sa vychádzalo zo Savoniusovho rotora.

SAVONIUSOV ROTOR - tento druh rotora bol základným druhom, z ktorého sa vychádzalo pri konštrukcii ďalších vertikálnych rotorov. Ide o pomalobežný rotor, ktorého výhoda je, že sa rozbieha pri nízkych rýchlostiach vetra, pretože zachytáva energiu z vetra z ktorejkoľvek strany celou plochou lopatiek. Lopatky sú konštruované v tvare písmena S, ako to je u rotorov používaných v praxi Obr.1. Tento rotor bol vylepšený realizáciou presahu lopatiek ich posunutím podľa Obr.1b čím došlo k vytvoreniu priepustného kanála , ktorý prispieva k zníženiu aerodynamického odporu a vzniku reaktívnej sily.

Na návrh modelov bolo potrebné nájsť čo najvhodnejší materiál, aby bolo možné pri

konštrukcii dosiahnuť požadovaný tvar a mechanické vlastnosti. Jednotlivé rotory boli rozmerovo identické. Na konštrukciu modelov veterných turbín bol použitý plast, ktorý má malú hmotnosť a dobrú pevnosť. Celkovo bolo skonštruovaných 16 veterných turbiniek rôznych tvarov. Všetky vychádzali z konštrukcie Savoniusovho rotora a jeho modifikácií týkajúcich sa presahov lopatiek a počtu stupňov. Priemery všetkých rotorov boli rovnaké 12 cm a ich výška bola 10 cm. Rozmery rotorov museli byť dodržané pre každý model, aby sa pri meraniach dali vzájomne porovnať ich vlastnosti .

Pre porovnanie vlastností boli skonštruované rotory bez presahu lopatiek, s jedno, dvoj a trojcentimetrovým presahom lopatiek. Výška lopatiek bola 10 cm a ich priemer bol závislý od konkrétneho presahu lopatiek. Rotor bez presahu lopatiek má priemer lopatky 6 cm, rotor s jednocentimetrovým presahom lopatiek má priemer lopatky 6,5 cm, rotor s dvojcentimetrový presahom lopatiek má priemer lopatky 7 cm a rotor s trojcentimetrový presahom lopatiek má priemer lopatky 7,5 cm. Presahom lopatiek sa získava väčšia efektívnosť využitia rotora, pretože rotor s presahom lopatiek sa skôr dokáže rozbehnúť a dosiahnuť vyššie otáčky ako klasický rotor bez presahu lopatiek. Spočíva to vo využití zákona akcie – reakcie.



Obr. 1 V reze Savoniusov rotor (vľavo) bez presahu lopatiek, (vpravo) s presahom lopatiek

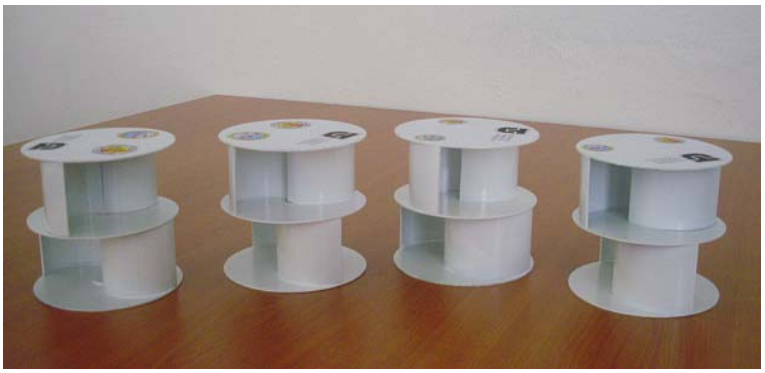
Pri konštrukcii všetkých rotorov sa vychádzalo z Obr. 1. Praktická realizácia je zrejmá z Obr. 2 Vzájomné odlišnosti sa týkali iba presahu lopatiek a počtu stupňov rotorov. Pri konštrukcii viacstupňových Savoniusových rotorov sa taktiež vychádzalo z Obr. 1, rozdiel bol len vo vzájomnom pootočení jednotlivých stupňov.



Obr. 2 Savoniusové rotory s presahom lopatiek 0-1-2-3 cm

VIACSTUPŇOVÉ ROTORY - Vzhľadom na to, že základný Savoniusov rotor dodáva počas otáčky iba dva pulzy výkonu a časť otáčky absolvuje iba zotrvačnosťou, boli skonštruované dvoj a trojstupňové rotory, ktoré sú vzájomne symetricky a súoso pootočené. Tieto rotory lepšie spracovávajú turbulentné prúdenie vzduchu. Sú to rotory zložené akoby z dvoch, alebo troch základných Savoniusových rotorov. Pri dvojstupňových Savoniusových rotoroch (Obr. 3) boli lopatky pootočené o 90°. Trojstupňové rotory (Obr. 4) majú jednotlivé stupne pootočené o 120°. Výhoda pootočenia jednotlivých stupňov je, že rotor dodáva viac pulzov výkonu na otáčku a má lepší záberový moment. Konštrukčne boli zhotovené rotory s rovnakými presahmi lopatiek ako v predchádzajúcom prípade. Vonkajšie rozmery rotora sú totožné avšak výška jednotlivých stupňov bola rozdelená podľa počtu stupňov. Keďže bolo

potrebné zhotoviť rotory s výškou 10 cm, tak tá sa pri viacstupňových rotoroch musela predeliť počtom stupňov. Výška lopatiek pre dvojstupňové rotory bola 5 cm a pre trojstupňové rotory 3,33 cm.



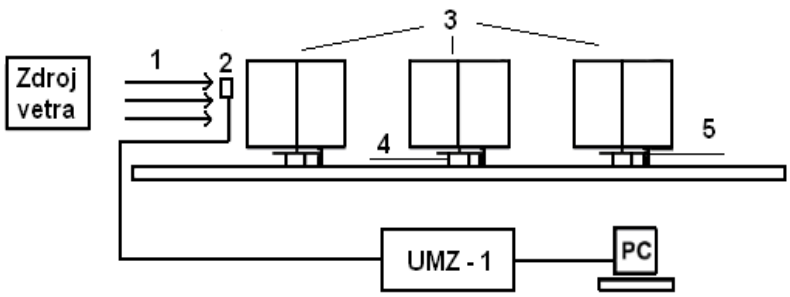
Obr. 3 Dvojstupňové Savoniusové rotory s presahom lopatiek 0-1-2-3 cm



Obr. 4 Trojstupňové Savoniusové rotory s presahom lopatiek 0-1-2-3 cm

III. KONŠTRUKCIA MERACEJ ZÁKLADNE

Meracia základňa bola skonštruovaná podľa blokovej schémy na Obr. 5 a tvorí ju drevený podstavec na ktorom sú umiestnené tri generátorčeky, na hriadeli ktorých je nasunutá kulisa s perforáciou na meranie otáčok rotorov, ktorá tvorí aj podstavec na umiestnenie rotorov pomocou magnetov. Tento spôsob predstavuje jednoduché a zároveň pevné držanie rotorov, ktoré sa dajú kedykoľvek ľahko odobrať z podstavca a vymeniť za iný rotor. Na meracej základni sú ďalej umiestnené snímače na meranie otáčok rotorov a elektronický anemometer na meranie rýchlosti prúdenia vzduchu. V spodnej časti meracej základne je umiestnená kabeláž od jednotlivých generátorčekov a snímačov otáčok, ktorá je vedená k plošnému spoju. Meracia základňa plní podmienku pre umiestnenie troch rotorov, ktoré je možné uviesť súčasne do prevádzky Obr. 6.



1 - prúdenie vzduchu, **2** - elektronický anemometer, **3** - vertikálne rotory, **4** - generátorček, **5** - snímač na meranie otáčok

Obr. 5 Schéma modelu veterného parku



Obr. 6 Konštrukčné riešenie meracej základne

Meranie prevádzkových parametrov bolo uskutočňované pomocou univerzálneho meracieho zariadenia UMZ-1. Toto zariadenie obsahuje meracie obvody a sériové porty RS232, ktoré slúžia na komunikáciu zariadenia s počítačom. Počet portov v zariadení je dva a z toho jeden slúži pre meranie otáčok rotorov a druhý na meranie napätia, ktoré je generované z generátorčekov. Meracie zariadenie obsahuje aj USB port, ktorým sa zabezpečuje napájanie zariadenia. Namerané hodnoty sú zobrazované pomocou PC.

MERANIE GENEROVANÉHO NAPATIA - modul pre meranie napätia poskytuje možnosť merať tri generátorčeky naraz, ktoré generujú napätie. Ďalšou možnosťou, ktorú tento modul poskytuje je meranie prúdenia vzduchu. Spolu je ním možné odmerať súčasne štyri hodnoty. Napájanie je riešené pomocou USB portu s hodnotou 5V. Impulzy od generátorčeka vedú cez LPT port do UMZ-1, odkiaľ sú ďalej cez sériový port RS232 vysielané do počítača, ktorý dané impulzy spracuje a zobrazí odmerané hodnoty pomocou programu PC Voltmeter.

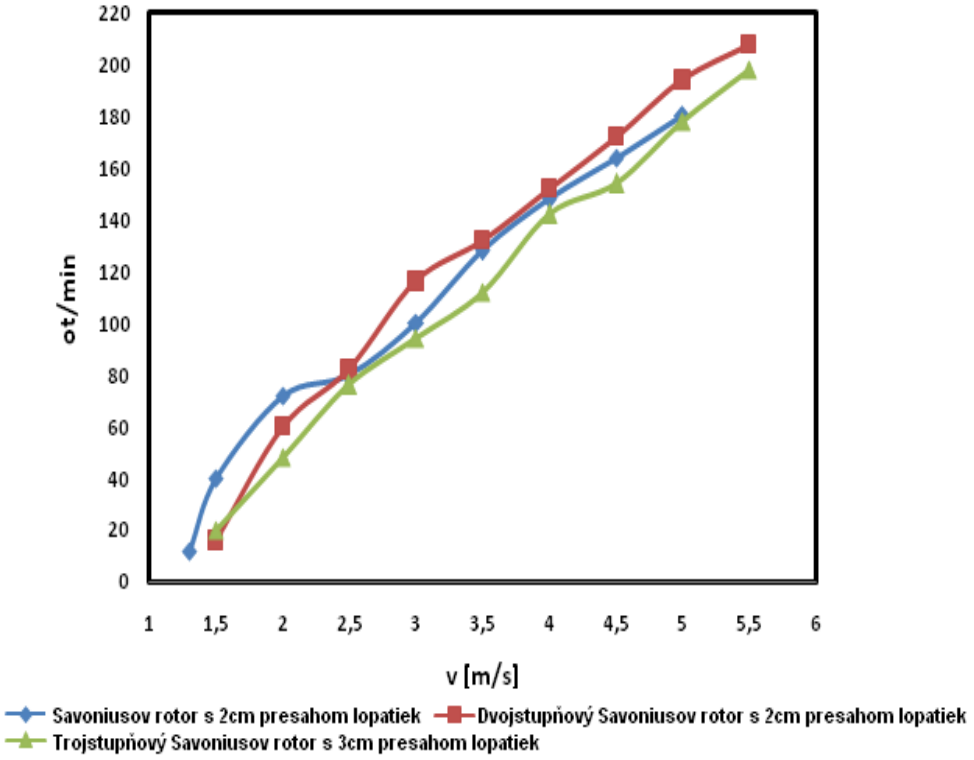
MERANIE OTÁČOK MODELU - na snímanie počtu otáčok bol použitý snímač otáčok, ktorý pracuje na princípe prerušenia svetla. Snímač otáčok vyšle impulz do UMZ-1 vtedy, keď sa preruší dopad svetla k fototranzistoru, ktorý vysiela fotodióda smerom k nemu. Prerušovanie svetla je pomocou otáčavej kulisy, ktorá je navrhnutá tak, že ak turbína urobí jednu otáčku, tak bude vyslaných 16 impulzov do UMZ-1. Týchto 16 impulzov je rovných jednej otáčke rotora. Snímače otáčok sú umiestnené na meracej základni a prostredníctvom LPT portu a kábla sú pripojené k meraciemu zariadeniu UMZ-1. Z meracieho zariadenia je impulz vyslaný cez port RS232 do počítača, ktorý dané impulzy vyhodnotí pomocou špeciálne vytvoreného programu PC Otáčkomer [7].

MERANIE RÝCHLOSTI PRÚDENIA VZDUCHU - pri meraní parametrov rotorov je potrebné vedieť aj rýchlosť prúdenia vzduchu. Keďže ide o modely veterných turbín, ich rozmery sú malé a meranie rýchlosti vetra anemometrami by nebolo možné, pretože ich rozmery sú veľké a výrazne by ovplyvnili prúdenie vzduchu smerom k turbíne. Kvôli tomuto problému bol skonštruovaný elektronický anemometer pracujúci na princípe zmeny odporu vlákna vplyvom teploty. Bol výhodný z hľadiska jeho malých rozmerov. Hodnoty rýchlosti prúdenia vzduchu sú zobrazované pomocou špediálne na tento účel vytvoreného programu PC voltmeter [7].

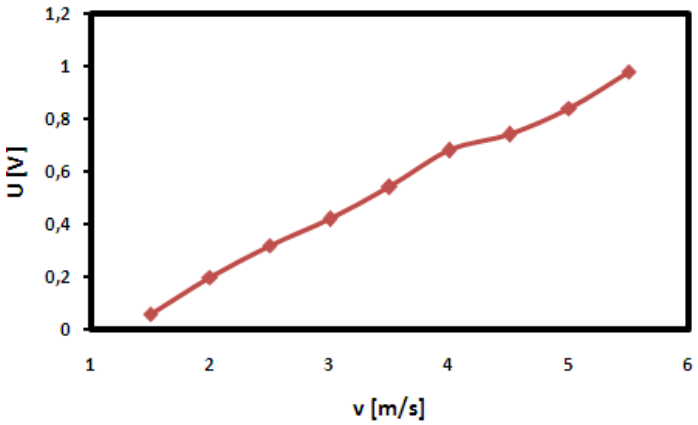
IV. Experimentálne Merania na Modeloch Veterných Turbín

Merania vlastností rotorov boli realizované pri rôznych rýchlostiach prúdenia vzduchu a z nameraných hodnôt boli vypracované grafické závislosti otáčok, napätia a prúdu v závislosti od rýchlosti prúdenia vzduchu. Namerané závislosti otáčok pre vybrané rotory sú na Obr. 7. Z grafickej závislosti na Obr. 7 je vidieť, že rozbeh rotorov sa začal už pri rýchlosti prúdenia vzduchu 1,3 m/s, čím sa potvrdilo, že tieto druhy rotorov sa rozbiehajú pri nízkych rýchlostiach.

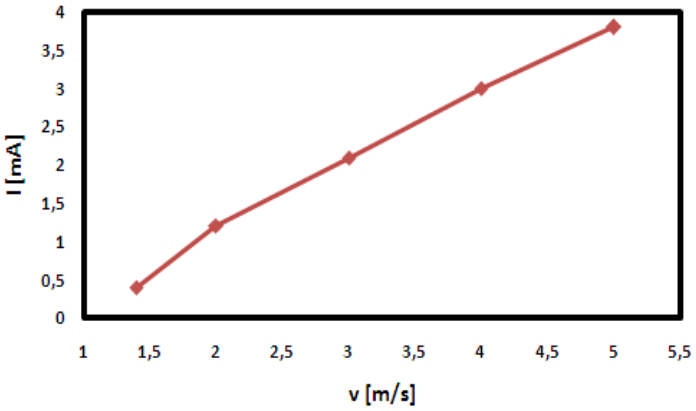
Z Obr. 7 je zrejmý aj vplyv počtu stupňov na linearizáciu charakteristiky v dôsledku väčšieho počtu pulzov výkonu počas jednej otáčky. Najlepšie výsledky mal z tohto pohľadu trojstupňový rotor s presahom 3 cm. Namerané závislosti výstupného napätia a prúdu pre trojstupňový rotor sú na Obr. 8. a Obr. 9.



Obr. 7 Závislosť otáčok vybraných rotorov od rýchlosti vetra



Obr. 8 Závislosť výstupného napätia od rýchlosti vetra



Obr. 9 Závislosť výstupného p1rúdu od rýchlosti vetra

Z množstva meraní bol vytvorený katalóg rotorov s popisom konštrukcie, nameranými hodnotami a závislosťami jednotlivých rotorov.

V. ZÁVER

V posledných rokoch inštalované výkony veterných elektrární rapídne stúpa, čo predstavuje vyššiu produkciu elektrickej energie. Princíp vo využívaní energie z vetra spočíva v premene kinetickej energie na mechanickú, ktorá dá do pohybu generátor a ten produkuje elektrickú energiu. V súčasnosti nachádzajú najväčšie uplatnenie rotory s horizontálnou osou otáčania, ktoré tvoria dve, alebo tri lopatky, ktoré z hľadiska aerodynamiky sú čo najlepšie natočené do smeru vetra. Konštrukcia takýchto rotorov nie je ani zďaleka tak jednoduchá ako sa na prvý pohľad môže zdať, pretože tieto rotory musia byť navrhované tak, aby ich využitie bolo čo najlepšie. Podstatne ľahšiu konštrukciu majú rotory s vertikálnou osou otáčania. Najznámejšie rotory z tejto koncepcie sú Savoniusov a Darrieusov rotor [5]. . Výhodou u týchto rotorov je, že nepotrebujú byť natáčané do smeru vetra a do pohybu sa dostávajú už pri nízkych rýchlostiach vetra. Nachádzajú dobré uplatnenie v oblastiach, kde nie je vietor stabilná rýchlosť a smer vetra.

Konštrukcia modelov veterných turbín spočívala v zostrojení modelu veterného parku. Konštrukčne boli zhotovené turbíny s vertikálnou osou otáčania, ktoré majú ľahšiu konštrukciu a nepotrebujú natáčanie do smeru vetra. Z nameraných hodnôt sa zistilo, že skonštruované rotory vyhovujú laboratórnym podmienkam a tvoria model veternej farmy. Rotory sa dostali do pohybu pri rýchlostiach vzdušného prúdu o hodnote od 1,2 – 2 m/s, čím sa potvrdilo, že sa rozbiehajú pri nízkych rýchlostiach vetra. Pri zvyšovaní rýchlosti vzdušného prúdu sa okrem otáčok zvyšovalo aj napätie z generátorčeka. Pre zistenie prúdov sme zapojili ku generátorčeku záťaž s hodnotou 200Ω, podľa ktorej sme generovaný prúd vypočítali.

Najlepšie výsledky vykazovali dvojstupňové a trojstupňové Savoniusové rotory, ktoré dobre spracovali vzniknuté turbulencie a dokázali by pracovať ešte s vyššími rýchlosťami vetra, ako sme namerali v laboratórnych podmienkach. Klasické Savoniusové rotory majú pomerne vysoký aerodynamický odpor. Čo sa prejavovalo hlavne pri vyšších rýchlostiach prúdenia vzduchu. Jednotlivé konštrukčné typy turbín vykazovali dobrú stabilitu v celom rozsahu rýchlosti prúdenia vzduchu, takže ich nebolo potrebné vyvažovať. V konečnom dôsledku tieto turbíny tvoria spolu s podstavcom model veternej farmy, ktorá môže slúžiť v ďalšom období na rôzne merania, alebo ako ukážka prevádzky veterných elektrární s vertikálnou osou otáčania.

Vertikálne rotory sa v súčasnosti často používajú na zabezpečenie elektriny v autonómnych systémoch v spolupráci s fotovoltickými panelmi čím sa dosahuje lepšia stabilita a spoľahlivosť dodávaného výkonu.

POĎAKOVANIE

Táto práca bola podporovaná Agentúrou na podporu výskumu a vývoja prostredníctvom finančnej podpory projektu č. APVV-0185-10

LITERATÚRA

[1] R. Rybár, D. Kudelas, G. Fischer, Alternatívne zdroje energie 3 **:** veterná energia / - 1. vyd - Košice : Edičné stredisko/AMS, - 2004. - 99 s. - ISBN 80-8073-144-6.

[2] I.Viťaz,M. Ďulik, Z. Dostál, *Prenos* dát meraných meteorologických veličín do počítača*. In.:Zborník z konferencie „ALER 2008“,* Liptovský Ján, 9.–10.10.2008, ISBN 978-80-8070-912-9, str. 79- 86.

[3] D. Kudelas, R. Rybár, M. Cehlár, Energia vetra **:** prírodné, technické a ekonomické podmienky jej využitia / - 1. vyd - Košice : FBERG TU, - 2009. - 215 s. - ISBN 978-80-553-0169-3.

[4] M, Taušová,, P. Tauš, J. Koščo, "Ekonomika ostrovnej FV elektrárne v podmienkach Slovenska". *Zborník prednášok ALER 2009 , 5. ročník, str.1-8,* Liptovský Ján 2009, ISBN 978-80-554-0099-0

[5] Z. Dostál, M. Bobek, J. Župa ,"Meranie globálneho slnečného žiarenia" *Zborník konferencie RESPECT, 2. ročník, Tatranská Lomnica, 3. – 5. 10. 2007, In: Acta Montanica Slovaca, roč. 13 (2008), číslo 3, str. 355-360, ISSN 1335-1788*

[6] P. Bača, "Ostrovní systémy: Problematika akumulace elektrické energie z FV do olověného akumulátoru". *In 30. NEKONVENČNÍ ZDROJE ELEKTRICKÉ ENERGIE. BRNO: ČES, 2009. s. 70-73. ISBN: 978-80-02-02164- 3.*

[7] Š. Hudák, S. Šimoňák, Programovacie techniky, *FEI TU Košice - 1. vyd - 2010. - 220 s*. - ISBN 978-80-553-0531-8

Využitie SVC na minimalizáciu činných strát v elektrizačnej sústave

Roman JAKUBČÁK, Ľubomír BEŇA

Katedra elektroenergetiky, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

roman.jakubcak@tuke.sk, lubomir.bena@tuke.sk

***Abstrakt* — Tento článok sa zaoberá možnosťou využitia SVC (Static var compensator) v elektrizačnej sústave (ES), so zameraním na zníženie činných strát, ku ktorým dochádza pri prenose výkonu v sieti. Súčasne sa sleduje vplyv SVC na zmenu napätí v jednotlivých uzloch siete. Činné straty je možné znížiť vhodným umiestnením SVC do sústavy s optimálnym nastavením jeho parametrov. Všetky uvedené výpočty boli realizované v programe MALTAB.**

***Kľúčové slová* — FACTS, SVC, genetické algoritmy**

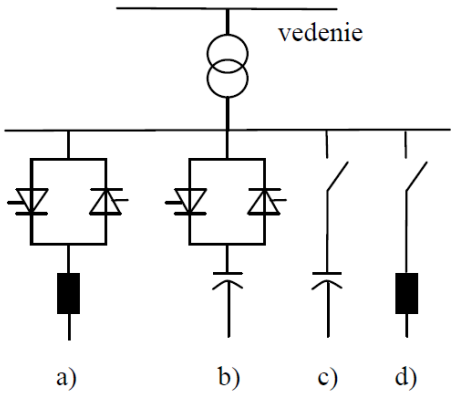
I. Úvod

Reguláciou jalového výkonu je možné zvýšiť napäťovú stabilitu ES a súčasne znížiť straty, ku ktorý dochádza pri prenose výkonu. Za týmto účelom je možné využiť zariadenia FACTS (Flexible Alternating Current Transmission System). Jedným z týchto zariadení je aj SVC, ktoré umožňuje riadiť jalový výkon v mieste jeho pripojenia, čím je schopný ovplyvniť napätie v danej oblasti.

II. Statický Kompenzátor Jalového Výkonu

Statický kompenzátor jalového výkonu (SVC) pozostáva z paralelne zapojeného statického zdroja alebo spotrebiča jalového výkonu, ktorého funkciou je regulovať určité parametre v sústave (najčastejšie je to napätie v uzle). Je to univerzálny názov pre tyristorom riadenú alebo spínanú tlmivku, kondenzátor alebo ich kombináciu [1].

Vo väčšine prípadov sa ako optimálne riešenie javí kombinácia tyristorovo spínaného kondenzátora s tyristorovo riadenou tlmivkou, ktorá poskytuje plynulú zmenu jalového výkonu cez celý riadiaci rozsah. Umožňuje pracovať v induktívnej aj kapacitnej oblasti [5].



Obr. 1 Statický kompenzátor jalového výkonu: a) tyristorom riadená/spínaná tlmivka, b) tyristorom spínaný kondenzátor, c) mechanicky spínaný kondenzátor, d) mechanicky spínaná tlmivka [3].

Model SVC pozostáva zo zdroja jalového výkonu, ktorý je zapojený do zvoleného uzla siete. Tento zdroj je schopný dodávať/odoberať jalový výkon do/zo siete, čím je schopný ovplyvňovať napätie v uzloch siete. Pri výpočtoch sa tento model prejavuje zmenou odoberaného, resp. dodávaného jalového výkonu v uzle, ktorom je SVC zapojené.

III. FORMULÁCIA ÚLOHY A VYUŽITIE GENETICKÝCH ALGORITMOV

Úlohou je nájsť najvhodnejšie umiestnenie SVC s konkrétnymi parametrami v ES, ktoré bude viesť k zníženiu činných strát v sieti. Tieto straty predstavujú rozdiel medzi generovaným a spotrebovaným činným výkonom v sieti. Bilancia činných výkonov v sieti je daná vzťahom:

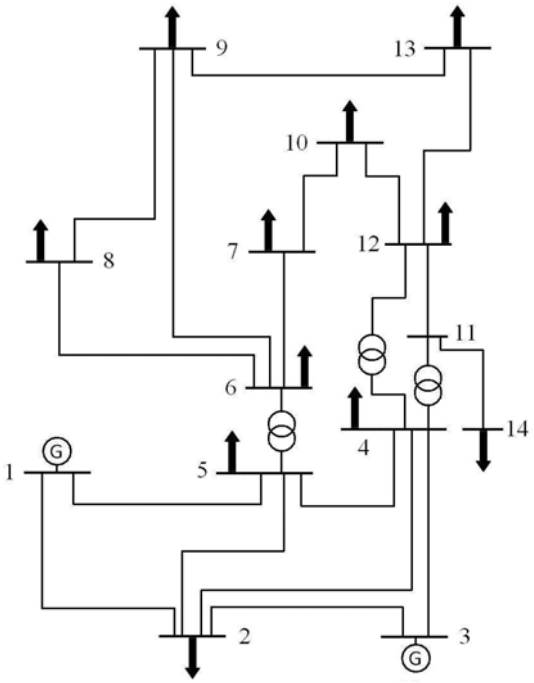
$$\sum P_G = \sum P_S + \sum \Delta P \tag{1}$$

Kde P_G predstavuje všetok činný výkon dodaný do siete, P_S odoberaný činný výkon a ΔP činné straty.

Na nájdenie najvhodnejšieho umiestnenia v sieti s vhodnými parametrami SVC bol použitý genetický algoritmus (GA). GA je vhodný na riešenie optimalizačných problémov, ktoré nie je vhodné riešiť pomocou štandardných optimalizačných metód, u ktorých je cieľová funkcia nespojitá, nediferencovateľná, náhodná, alebo silne nelineárna. Proces výberu u GA je založený na prirodzenom výbere a je riadený biologickou evolúciou. Táto evolúcia spôsobuje neustálu modifikáciu populácií riešení a cez úspešné generácie riešení sa postupne približuje k optimálnemu riešeniu. Na vytvorenie novej generácie GA využíva viacero nástrojov: selekcia, kríženie a mutácia. Proces výpočtu sa zastaví ak je dosiahnutá niektorá z podmienok na ukončenie výpočtu [2]. Týchto podmienok existuje viacero, v tomto konkrétnom prípade to je podmienka zastaviť výpočet ak GA už nie je schopný nájsť menšiu hodnotu cieľovej funkcie oproti tej, ktorú už dosiahol.

Cieľovú funkciu prestavuje suma činných strát, ktorej hodnota sa mení v závislosti od zmeny jalového výkonu v jednotlivých uzloch. Hľadané parametre sú:

- miesto zapojenia SVC do siete (uvažovaná sieť je na obr. 2).
- parametre SVC, t.j. hodnota dodávaného/odoberaného jalového výkonu.

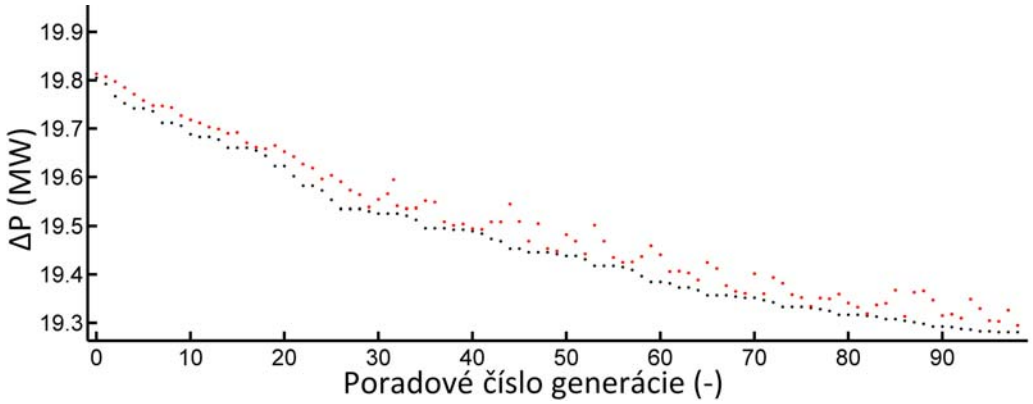


Obr. 2 14-uzlová sieť.

Nakoľko SVC slúži na reguláciu napätia v sieti, súčasne so zmenou činných strát je sledovaný aj vplyv na zmenu napätí v jednotlivých uzloch siete. Podmienkou je, aby sa napätie nezmenilo mimo povolený rozsah ± 5 % na napäťovej hladine 400 kV a ± 10 % na napäťovej hladine 110 kV.

Výpočtom pomocou GA bolo určené ako najvhodnejšie miesto z hľadiska minimalizácie strát v sieti uzol č. 9, do ktorého sa má dodávať jalový výkon o veľkosti 42,8 MVAr. Priebeh hľadania výpočtu pomocou GA je zobrazený na obr. 3. Čierne značky reprezentujú najmenšiu hodnotu cieľovej funkcie v danej generácii, červené značky reprezentujú priemernú hodnotu pre celú generáciu.

V tabuľke 1a 2 sú uvedené jednotlivé hodnoty výroby a spotreby v uzloch (kladná hodnota znamená výrobu a záporná hodnota spotrebu výkonu v uzle), zmeny napätí v jednotlivých uzloch a zmeny činných strát v jednotlivých vedeniach bez a s použitím SVC.



Obr. 3 Priebeh zmeny činných strát v závislosti od generácie riešení.

Tabuľka 1
Vplyv použitia SVC na napätia v uzloch siete.

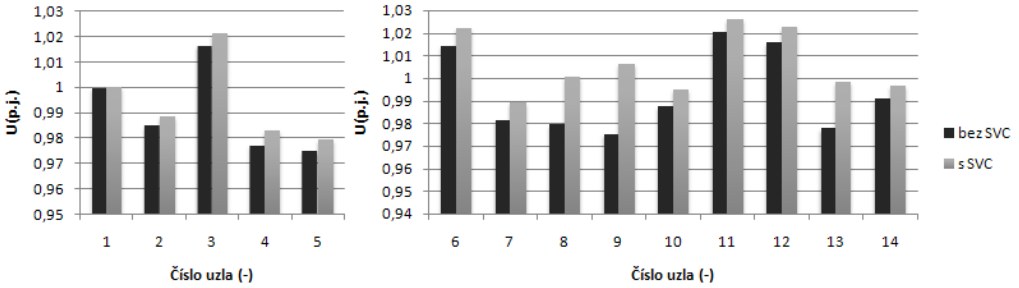
číslo uzla	výroba/ spotreba *P* (MW)	výroba/ spotreba *Q* (MVAr)	bez SVC *U* (kV)	s SVC *U* (kV)
1	1139,824	137,498	400,0	400,0
2	-400	-175	394,0	395,5
3	400	250	406,7	408,5
4	-400	-175	390,9	393,1
5	-400	-175	390,0	391,9
6	-40	-15	111,6	112,4
7	-40	-15	108,0	108,8
8	-40	-15	107,7	110,0
9	-40	-15	107,3	110,7
10	-40	-15	108,6	109,5
11	0	0	112,2	112,9
12	-40	-15	111,7	112,5
13	-40	-15	107,5	109,8
14	-40	-15	109,0	109,7

Tabuľka 2
Vplyv použitia SVC na činné straty v jednotlivých vedeniach siete.

číslo vedenia	z uzla	do uzla	bez SVC Δ*P* (MW)	s SVC Δ*P* (MW)
1	1	2	4,095	4,058
2	5	2	0,067	0,057
3	5	4	0,383	0,381
4	4	2	0,392	0,382
5	4	3	2,120	2,075
6	2	3	0,583	0,603
7	1	5	6,624	6,512
8	7	6	0,965	0,959
9	8	6	1,228	1,092
10	9	8	0,043	0,113
11	9	6	0,924	0,873
12	10	7	0,020	0,018
13	12	11	0,027	0,023
14	10	12	0,715	0,697
15	13	12	0,851	0,642
16	13	9	0,020	0,033
17	11	15	0,767	0,757
celkové Δ*P* (MW)			19,82	19,28

Po zapojení SVC do uzla č. 9 sa straty v sieti znížili z hodnoty 19,82 MW na 19,28 MW, čo je zníženie o 2,76 %. Je potrebné poznamenať, že činné straty klesli v oboch napäťových

hladinách, t.j. na napäťovej hladine 400 kV aj 110 kV. Priebeh zmeny napätí s a bez použitia SVC na napäťovej hladine 400 kV (vľavo) a 110 kV (vpravo) je uvedený na obr. 4 v pomerných jednotkách.



Obr. 4 Priebeh zmien napätí v uzloch siete bez a s použitím SVC.

IV. ZÁVER

V článku bolo preukázané, ako pomocou využitia SVC je možné meniť nie len napätia v uzloch siete, ale aj znížiť činné straty. Za týmto účelom je potrebné nájsť vhodné umiestnenie a nastavenie SVC v ES. Potrebné simulácie boli robené v programe MATLAB a v toolboxe Power System Analysis Toolbox (PSAT).

POĎAKOVANIE

Táto práca vznikla za podpory Agentúry na podporu výskumu a vývoja v rámci projektu APVV-0385-07 a SK-BG- 0010-08, a taktiež Vedeckej grantovej agentúry Ministerstva školstva Slovenskej republiky a Slovenskej akadémie vied v rámci projektu VEGA-1/0166/10.

LITERATÚRA

[1] GRIGER, V., GRAMBLIČKA, M., NOVÁK, M., POKORNÝ, M.: Prevádzka, riadenie a kontrola prepojenej elektrizačnej sústavy. Žilina: Vydavateľstvo ŽU, 2001, 236 s. ISBN 80-7100-823-0.

[2] Genetic algorithm and direct search toolbox, Dostupné na internete: http://www.mathworks.com/help/releases/R13sp2/pdf_doc/gads/gads_tb.pdf

[3] KOLCUN, M., BEŇA, Ľ., MÉSZÁROS, A., RUSNÁK, J.: Riešenie problémov v riadení prevádzky elektrizačných sústav s využitím FACTS zariadení. In: Elektroenergetika 2005 : Zborník z 3. vedeckého sympózia : 21.-23. september 2005, Stará Lesná, Vysoké Tatry. Košice : FEI-TU v Košiciach, 2005. 23 s. ISBN 80-8073-305-8.

[4] PISICA, I.; BULAC, C.; EREMIA, M.: Optimal SVC placement in electric power systems using a genetic algorithms based method. In.: Intelligent System Applications to Power Systems, 2009, IEEE Bucharest Power Tech Conference, June 28th - July 2nd, Bucharest, Romania. E-ISBN : 978-1-4244-2235-7. Dostupné na internete: http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=5281841&url=http%3A%2F%2Fieeexplore.ieee.org%2Fiel5%2F5271856%2F5281781%2F05281841.pdf%3Farnumber%3D5281841

[5] JOHNS, A. T., TER-GAZARIAN, A., WARNE, F.: "Flexible ac transmission systems (FACTS)," The Institution of Electrical Engineeres, 1999, 592 pp, ISBN 0-85296-771-3

Návrh a konštrukcia ovládacej jednotky pre polohovacie zariadenie

František JUSKO, Ján TKÁČ

Department of Electric Power Engineering, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Slovak Republic

frantisek.jusko@tuke.sk, jan.tkac@tuke.sk

***Abstrakt* — The present work deals with the use of solar energy, focusing on its direct conversion into electricity. This conversion takes place by means of photoelectric cells. Given their relatively low efficiency to seek ways by which they could be energy profit increased. One of the ways to achieve higher energy profit is positioning photovoltaic solar cells direct to Sun. The aim of this work was to design and construction a control unit for laboratory positioning device. The basis for its construction was previously constructed a positioning device operated only manually, without the possibility of collaboration with the PC. Engineered control device was designed to allow the attitude control via a PC, or automatically calculated based on the coordinates of the Sun. Verification measurements confirmed the functionality of a sufficient positioning accuracy towards the sun.**

***Kľúčové slová* — slnečná energia, polohovacie zariadenie, ovládacia jednotka, konštrukcia**

I. Úvod

Polohovacie zariadenia v solárnej technike sú jednou z možností ako zvýšiť energeticky zisk fotoelektrických článkov, kolektorov. V solárnych elektrárňach a koncentrovaných solárnych zariadeniach sú polohovacie zariadenia nevyhnutnosťou.

Prvou požiadavkou pri konštrukcií tejto polohovacej jednotky bolo, aby sa laboratórne polohovacie zariadenie dalo ovládať pomocou PC. Hlbšou analýzou tejto požiadavky bola ovládacia jednotka doplnená o ďalšie požiadavky:

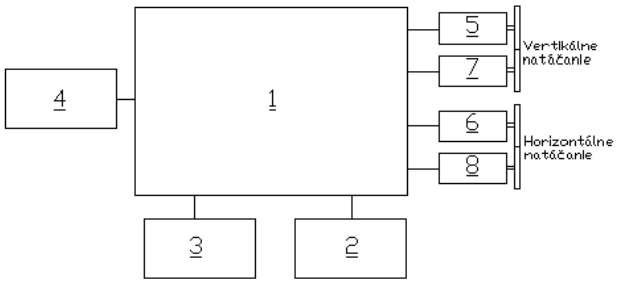
- automatická prevádzka,
- manuálne nastavenie polohy,
- spolupráca s osobným počítačom,
- jednoduchosť zapojenia,
- napájanie bezpečným napätím,
- možnosť ďalšieho rozšírenia.

Zo systémov na sledovanie polohy Slnka sme si zvolili systém s otvorenou slučkou, riadený a ovládaný mikrokontrolérom [1].

II. Konštrukcia Ovládacej jednotky polohovacieho Zariadenia

Návrh ovládania polohovacieho zariadenia vychádzali z požiadaviek na jeho funkciu pri laboratórnych meraniach solárnych zariadení, a preto obsahuje niektoré činnosti ktoré nie sú inštalované v komerčne vyrábaných polohovacích systémoch.

Ovládacie zariadenie bola skonštruované podľa blokovej schémy na obr. 1.



Obr. 1 Bloková schéma ovládacej jednotky.

Jeho základom je ovládacia jednotka OVJF3 **1** obsahujúca programové vybavenie pre zabezpečenie správnej činnosti polohovacieho zariadenia. Táto jednotka je napájaná z napájacieho zdroja15 – 25V **2**, prípadne z akumulátora12V **3**, ktorý umožňuje autonómnu, alebo záložnú prevádzku polohovacieho zariadenia. Spôsob napájania umožňuje vyššiu spoľahlivosť polohovacieho zariadenia a to aj v spolupráci s využívaním energie fotovoltaického zdroja. Ovládacia jednotka prostredníctvom spínacích obvodov ovláda servopohon **5** pre nastavenie vertikálnej polohy montážnej plošiny a servopohon **6** pre nastavenie horizontálnej polohy montážnej plošiny. Servopohony **5** a **6** sú vybavené odporovými vysielačmi **7** a **8** ktoré poskytujú informácie o okamžitej polohe montážnej plošiny, na základe ktorých sa realizuje polohovanie. Vysielače sú zapojené ako deliče napätia. Vstupné napätie deliča je 5 V a výstupné napätie závisí od uhla natočenia vysielača.

K ovládacej jednotke **1** je možné pripojiť prostredníctvom USB rozhrania osobný počítač **4**, pomocou ktorého môžeme ovládať polohovacie zariadenie a jeho nastavenie do optimálnej , alebo požadovanej polohy.

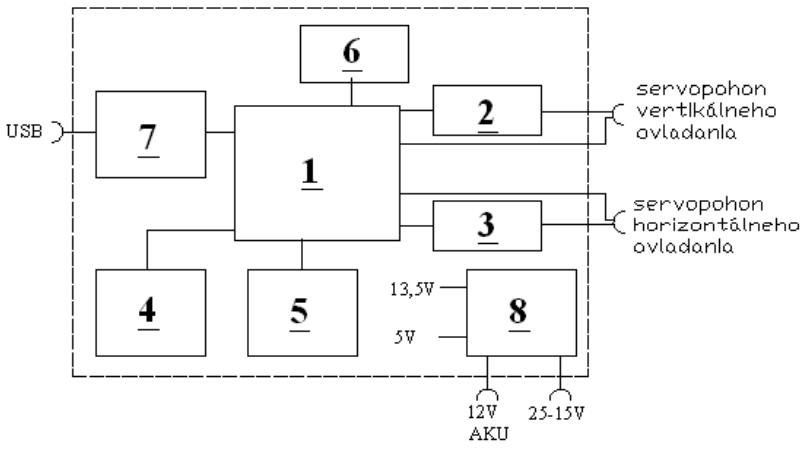
Konštrukčné riešenie polohovacieho zariadenia s ovládacou jednotkou je na Obr.2.



Obr. 2 Konštrukčné usporiadanie polohovacieho zariadenia pri meraní

Konštrukčné riešenie ovládacej jednotky OVJF3

Ovládacia jednotka bola skonštruovaná podľa blokovej schémy na Obr. 3.



Obr. 1 Bloková schéma ovládacej jednotky OVJF3

Základom ovládacej jednotky je mikrokontrolér **1** obsahujúci programové vybavenie pre riadenie činnosti polohovacieho zariadenia. Zabezpečuje ovládanie smeru pohybu montážnej plošiny prostredníctvom prepínača smeru vertikálnej polohy **2** a horizontálnej polohy **3**. Mikrokontrolér **1** dostáva informáciu o polohe montážnej plošiny z odporových vysielačov polohy servopohonov. K zabezpečeniu správneho výpočtu polohy Slnka pri automatickom polohovaní je táto jednotka vybavená obvodom reálneho času **4**. Jednotka komunikuje s PC pomocou prevodníku USB/Seria **7**.



Obr. 4 Čelný panel ovládacej jednotky

Všetky obvody riadiacej jednotky sú umiestnené v plastovej skrinke o rozmeroch 145x200x600 mm. Tá je umiestnená na konštrukcii dvojosého polohovacieho zariadenia. Z pravej strany sú umiestnené konektory na pripojenie servopohonov z ľavej strany je umiestnený konektor USB a jednokolíková zásuvka na pripojenie napájania. Na prednej strane skrinky je umiestnený priezor pre displej a ovládacie tlačidlá. Elektronické obvody sú osadené na troch jednovrstvových doskách plošného spoja.

Na ovládanie jednotky, pre manuálne polohovania trackera, na nastavovanie vstupných údajov a prepínanie módov prevádzky je na čelnom paneli umiestnená klávesnica **5** obsahujúca šesť ovládacích tlačidiel s týmito funkciami:

- ▲ pohyb hore, posun kurzora hore, nárast hodnoty,
- ▼ pohyb dole, posun kurzora dole, pokles hodnoty,
- ◄ natáčanie vľavo, posun kurzora doľava,
- ► natáčanie vpravo, posun kurzora doprava,
- ENTER posun v menu na vyššiu úroveň, potvrdenie nastavenej hodnoty,
- ESC návrat späť.

K zobrazeniu údajov o zvolenom pracovnom režime, UTC čase, polohy montážnej plošiny, kalibračných údajov, potrebných na ovládanie tejto jednotky bol použitý bodový displej z tekutých kryštálov GLCD **6**, s rozlíšením 128x64 bodov. Je zapojený podľa schémy zapojenia **Chyba! Nenašiel sa žiaden zdroj odkazov.** Pre zníženie spotreby energie bol displej doplnený o obvod zapnutia a vypnutia podsvietenia.

Spoluprácu ovládacej jednotky s osobným počítačom zabezpečuje USB rozhranie **7**, prostredníctvom ktorého sa realizuje:

- inštalácia ovládacieho programu do mikrokontroléra,
- ovládanie polohovacieho zariadenia pomocou PC,
- posielanie údajov o vypočítanej polohe Slnka, údajov o polohe montážnej plošiny do PC.

ISBN 978-80-553-0890-6 © 2012 FEI TUKE

Vnútorné napájanie jednotlivých blokov ovládacej jednotky zabezpečuje napájací zdroj **8**, ktorý upravuje výstupné napätie vonkajšieho zdroja (15 – 25 V) na 5 a 13,5 V. Napätím 5 V napájame bloky:

- mikrokontrolér,
- obvod reálneho času,
- ovládacie tlačidlá,
- zobrazovacia jednotka GLCD,
- odporové vysielače servopohonov.

Napätie 13,5 V napája obvody prepínania smeru pohybu servopohonov **2** ,**3** čo je zabezpečené mostíkovým prepínačom (H-bridge).

III. Reálne Chyby Polohovacieho Zariadenia

Skutočnú presnosť polohovania sme overili ovládaním dvojosého polohovacieho zariadenia [2]. Ako sme spomenuli v teoretickej časti tejto práce, hlavnými príčinami nepresnosti pri systéme ovládania s otvorenou slučkou sú:

- chyby matematického výpočtu,
- chyby spôsobené nesprávnou inštaláciou,
- mechanické chyby konštrukcie.

Aj napriek tomu, že ovládacia jednotka vypočíta polohu Slnka s priemernou chybou v azimute 0,0012°, v elevácií 0,03992°, reálne hodnoty presnosti polohovania sa môžu značne odlišovať [3]. [5].

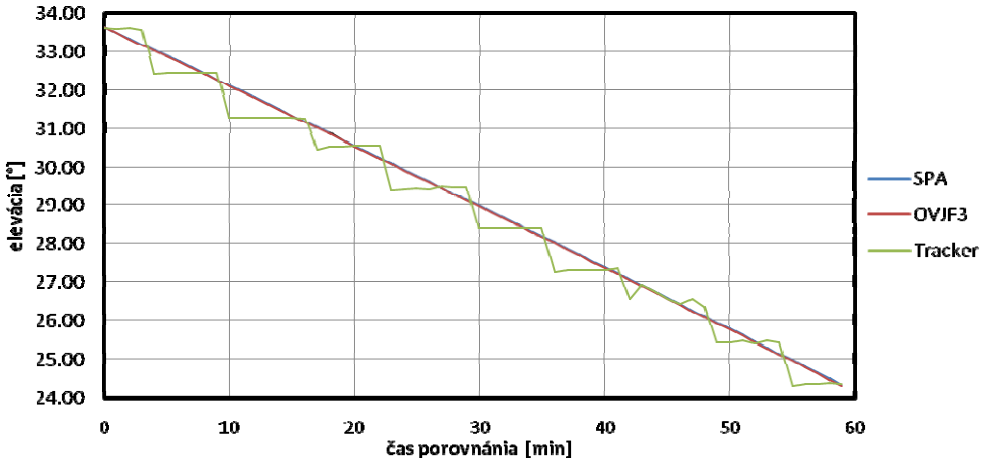
Inštalácia polohovacieho zariadenia

Polohovacie zariadenie, ktoré ovládala naša polohovacia jednotka, bolo vyrobené ako učebná pomôcka pre všeobecné merania v solárnej technike. Jeho konštrukcia je uspôsobená tak, aby ho bolo možné ľahko premiestniť na miesto merania. Mobilita tohto zariadenia so sebou prináša rad problémov pri jeho inštalácii:

- pred každým meraním je potrebné najprv zorientovať polohovacie zariadenie do základnej polohy, po umiestnení na miesto merania sa musí vykonať nivelácia v priečnom a pozdĺžnom smere
- problémy s určením zemepisných svetových strán (buzola je nepoužiteľná),

Chyby, ktoré vznikajú pri inštaláciách polohovacích zariadení patria k najčastejším a zároveň majú na svedomí aj najväčšiu mieru nepresnosti.

Pomocou GPS súradníc bodu merania a GPS súradníc okolitých stavieb sme určili orientáciu svetových strán. Nastavili sme azimutovú os natáčania do vertikálneho smeru, skalibrovali azimut a eleváciu a nastavili UTC čas (v našom prípade miestny čas (-)2 hod.). Týmto bolo polohovacie zariadenie pripravené na svoju činnosť.



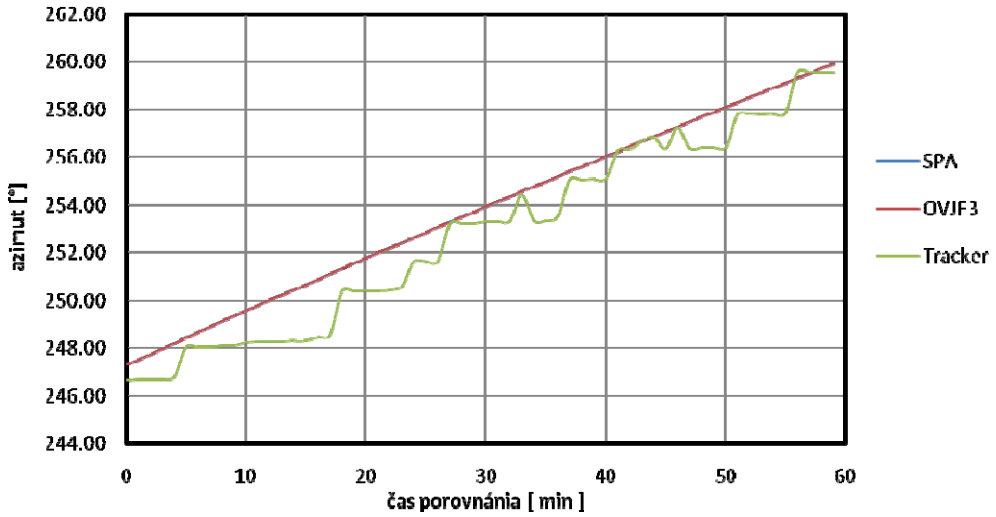
Obr. 5 Grafické znázornenie elevácie, vypočítanej od odchýlky

Z nameraných výsledkov vyplýva , že odchýlka v polohovaní elevácie do 15.00 hod UTC neprekročila jeden stupeň. Po 15.00 hod. sa začala prejavovať chyba matematického výpočtu spojená s lomom slnečných lúčov a odchýlka sa začala zvyšovať. Odchýlka polohovania

v azimute bola počas celého merania väčšia než odchýlka v elevácii. Pohybovala sa v rozsahu (-)0,4°– (+)2,5° do 15.50 hod. Po 15.50 hod. sa odchýlka začala zvyšovať z dôvodu, že polohovacie zariadenie bolo v krajnej polohe svojho pohybu a odchýlka mala už len narastajúci charakter.

Presnosť metódy určenia odchýlky azimutu a elevácie sme si overili grafickou metódou pomocou SPA algoritmu. Odchýlku sme odčítali od matematického výsledku algoritmu, čím by sme mali dostať reálnu polohu montážnej plochy v závislosti od času. Na obr. 5 a obr. 6 sú zobrazené vypočítané hodnoty azimutu a elevácie montážnej plochy, vypočítané podľa odchýlok v čase od 14.00 hod. do 14.59 hod. UTC.

Odchýlka azimutu a elevácie bola spôsobená nedokonalosťou polohovacieho zariadenia a hlavne jeho vysielača polohy.



Obr 6. Grafické znázornenie nastavenia azimutu montážnej plochy, podľa odchýlky

IV. Meranie Spotrebovanej Energie

Častou pripomienkou k využívaniu polohovacích zariadení je množstvo elektriny, potrebnej na zabezpečenie ich funkcie. Spotreba elektriny je závislá od požadovanej presnosti polohovania a času činnosti servopohonov, ktoré majú podstatný podiel na celkovej spotrebe polohovacieho zariadenia. Namerane hodnoty spotreby ovládacej jednotky, vrátane servopohonov sú uvedené v tabuľke 1.

Tabuľka 1
Namerané hodnoty odberu polohovacieho zariadenia .

	Ovládacia jednotka	**Polohovanie smerom hore**	**Polohovanie smerom dole**	**Polohovanie smerom vpravo**	**Polohovanie smerom vľavo**
Napätie [V]	16	16	16	16	16
Prúd [mA]	168	353	285	284	251

Z nameraných výsledkov vyplýva, že spotreba elektriny ovládacej jednotky pri nečinnosti servopohonov bola 168 mA. Z rozboru spotreby vyplýva, že z tejto spotreby podstatnú časť spôsobujú (cca 100 mA), odporové vysielače OV 100.s pomerne nízkym odporom (100 Ω).

V. Záver

Využívanie slnečnej energie sa v súčasnosti javí ako perspektívna náhrada konvenčných zdrojov energií. Účinnosť premeny slnečného žiarenia na tepelnú energiu je zvládnutá s vysokou efektivitou. Avšak účinnosť priamej premeny slnečného žiarenia na elektrickú energiu nedosahuje požadovanú ekonomickú rentabilitu.

Možnosti zvyšovania účinnosti tejto premeny sú vo vývoji nových materiálov, použitých na výrobu fotoelektrických článkov. Ďalšou možnosťou je zapájanie koncentrovaných solárnych zariadení, s ktorými sa dosahujú najvyššie účinnosti tejto premeny. Polohovacie zariadenia sú

nevyhnutnou súčasťou týchto zariadení. Obstarávacie náklady polohovacích zariadení predstavujú značnú časť nákladov pri výstavbe solárnych zariadení.

Cieľom našej práce bolo navrhnúť a prakticky overiť fungovanie jednoduchej, nenáročnej ovládacej jednotky, ktorá ovláda dvojosé polohovacie zariadenie. Polohovacie zariadenie ovládané skonštruovanou ovládacou jednotkou vykazovalo dostatočnú presnosť polohovania ako v automatickom režime, tak aj v režime ovládania pomocou PC a v manuálnom režime. Reálne odchýlky polohovania do doby dosiahnutia krajných polôh nepresahovali (+)2,57° a (-) 0,23° v azimute a v elevácii (+)1,46° a (-)0,43°.

Polohovacie zariadenie a jeho ovládanie je ešte možné v budúcnosti vylepšovať. Ďalšie zvyšovanie presnosti by bolo možné dosiahnuť inštaláciou presnejších snímačov nastavenej polohy. Takáto úprava by umožnila aj podstatné zníženie spotreby elektrickej energie. Z hľadiska mechanickej konštrukcie je potrebné venovať pozornosť odstráneniu vôli prevodových mechanizmov. Prínosom by bola aj úprava programu pre presnejší výpočet v čase východu a západu Slnka [2]. [3]. [4]. Veľmi vhodné by bolo doplnenie ovládacej jednotky o GPS navigáciu čím by sa zjednodušila a spresnila inštalácia polohovacieho zariadenia.

POĎAKOVANIE

Táto práca bola podporovaná Agentúrou na podporu výskumu a vývoja prostredníctvom finančnej podpory projektu č. APVV-0185-10.

LITERATÚRA

[1] R.Kittler, J.MIKLER, Základy využívania slnečného žiarenia., VEDA, Bratislava 1986. 150 s. 71-013-86

[2] Kok-Keong Chong a kol.: Integration On-Axis General Sun-Tracking Formula in the Algorithmofan Open-Loop Sun-Tracking System. [online]. [citované. 2012-04-09]. Dostupné na internete: http://www.mdpi.com/search/?q=kok&view=default&sortBy=relevance&pageCount=10&pageNo=2

[3] RUBIO, F.R. a kol.: Applicationof new control strategy for sun tracking. [online]. [citované. 2012-04-07]. Dostupné na internete:http://www.esi2.us.es/~rubio/ECM_07_Solar.pdf

[4] KOLCUN, M. a kol.: Elektrárne. Košice: Technická univerzita v Košiciach, 2006. 453 s. ISBN 80-8073-704-5.

[5] REDA, I. – ANDREAS, A.: Solar Position Algorithm for Solar Radiation Applications. [online]. [citované. 2012-04-13]. Dostupné na internete:http://oretano.iele-ab.uclm.es/~carrion/Bibliografia/Solar%20Position%20Algorithm%20for%20Solar%20Radiation%20Applications.pdf

Transformer Oil Based Ferrofluid From View of i(t) Characteristics

Jozef KIRÁLY, Martin GERMAN-SOBEK, Roman CIMBALA

Department of Electrical Power Engineering, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Slovak Republic

jozef.kiraly@tuke.sk, martin.german-sobek@tuke.sk, roman.cimbala@tuke.sk

***Abstract* — The article deals with i(t) characteristics of magnetic fluids based on a transformer oil and about influence of direction of acting static magnetic field. Thus measurement is dependent from orientation of actuating magnetic field. This behavior mainly shows the effect of nanoparticle structuring during exposition of specimen by magnetic field. It is also showed in article influence of nanoparticle concentration contained in fluid.**

***Keywords* — IRC analysis, ferrofluid, nanoparticles, transformer oil**

I. INTRODUCTION

Ferrofluid is liquid, which consist from carrier fluid and ferromagnetic nanoparticles coated by surface active substance. Respect particles size, ferrofluids are colloidal liquids. Common size of magnetite nanoparticles is from 2 to 20 nm. As mentioned, these particles are coated by surface active substance – surfactant. In our case takes the role of surfactant oleic acid, which is monounsaturated omega-9 fatty acid. Molecules of oleic acid are made from polar hydrophilic head and long unipolar hydrophobic chain. Main function of surfactant is to prevent from clustering. Particles which were created from many nanoparticles would be too heavy to stay dispersed in volume of liquid by Brownian motion. In case of usage ferrofluid as replacement of transformer oil is important to note that the settled particles can cause inhomogeneity in power transformer insulation system.

The wide possibilities of ferrofluid application are based on fact, that these fluids can be positioned inside technical device by acting external magnetic field. Most common use of ferrofluids today is as airtight and lubricating seals in rotary shafts. This is based on fact, that gradient of magnetic field keeps the magnetic fluid in correct position, even in case of pressure differences between the two separated compartments [1]

II. FERROFLUID AS A REPLACEMENT OF COMMON TRANSFORMER OIL

The transformer oil performs two functions in power transformers. Insulating, thus transformer oil prevents the flow of current between separate components as windings, core, vessel etc. Second function is usage as cooling medium, consequently conduct heat out of windings and core to walls of vessel. Subsequently is this heat dissipated into surrounding of vessel [1].

Another point of view is usage of transformer oil as impregnation of oil – paper insulation system. It should be noted, that the size of nanoparticles of Fe_3O_4 isn't limited only by presence of phenomenon of particles coupling caused by negotiation of Van der Waals force, but also with problem of penetration of paper in these systems. Density of these fluids must be sufficient low, because too dense magnetic fluid will not get into every layer of oil – paper insulation system.

Positive influence of application magnetic fluid instead of conventional transformer oil is lower temperature of windings and higher capacity to sustain overvoltage, and present better resistance at degradation in time due to humidity compared to classic oil. This fact brings smaller dimensions of power transformers and less amount of cooling liquid.

The better heat transfer from windings is caused by presence of phenomenon named thermomagnetic conductivity. Condition for this phenomenon is presence of temperature gradient and magnetic field. In case of usage in power transformer are conditions for thermomagnetic convection satisfied. The source of temperature gradient is heat produced by

AC current in windings. Thus heat generated by winding and core of transformer creates in area inhomogeneous magnetic field temperature gradient [2].

III. Analysis of I(T) characteristics

One of the non-destructive methods of diagnostics of high voltage insulation system in general is dielectric spectroscopy. This method is based on measurement of current in time and frequency domain. Stability of microstructure and composition of the insulating material is changing due to degradation processes. As a result of these changes is changing behavior of insulation material from view of polarization processes. Based on this fact, is possible to characterize degree of degradation of specimen. Our sample can be mentioned as a capacitor with pair of electrodes and dielectrics between them. Thus after connecting voltage to our specimen current starts flow thru dielectric. In case of ideal dielectric the current hasn't real component. But in practice is real component of current caused by processes of polarization and conductivity. These processes are dependent on kind of dielectric in which they are occurs [3]. Based on existence independent Debye's polarization processes and Maxwell – Wagner equivalent model is possible to formulate total polarization current as:

$$i_c = \frac{U}{R_i} + \sum_{i=1}^{n} I_{m_i} . e^{-\frac{t}{\tau_i}} \quad \text{(A;V, Ω,A,s)} \tag{1}$$

This equation can be used if for measurement is used electric field with intensity approx. 10^5 Vm^{-1}. The macroscopic view is descripted by current i_c which represents mutually independent polarization processes in magnetic fluid. U is applied voltage on the specimen, R_i is DC insulation resistance after infinite long time, I_{mi} is amplitude of i – th component of Debye's elementary current, τ_i is relaxation time constant of i – th component [4].

Usage of dielectric spectroscopy for examination of structuralization processes and processes of creating clusters of Fe_3O_4 nanoparticles is suitable because of measurement time provides sufficient view of ongoing polarization processes. It's possible to recognize these processes by increase of polarization current. It should be noted, that magnitude of current is changing by acting of magnetic field with different vector orientation. Also by usage of this method for diagnostic magnetic fluids it's necessary to apply external magnetic field before start of measurement, because structuralization of particles takes a short time (approx. 60 s) [5].

IV. Experiment

For measuring of test specimens is necessary to prepare test bench which allows various setup of acting external stationary magnetic field. In our case, is external magnetic field provided by pair of NdFeB magnets with magnetic induction 40 mT. Specimen was place in vessel with volume approx. 100 ml which is made from PTFE. For measurement of i(t) characteristics was used electrode system consisting of pair of parallel cylindrical electrodes with diameter 20 mm made from copper. The gap between electrodes was set to 1 mm because of creating electric field with sufficient intensity E.

By reason of comparison is indispensable use for measurement ferrofluids with different concentration of magnetic nanoparticles. For confirmation of ongoing phenomena we use fluids with different concentration based on same carrier fluid.

Therefore for measurement we use two kinds of specimens. Both of them are made on common inhibited transformer oil as carrier fluid. Likewise as surfactant is used oleic acid. For measurement of anisotropy are used two different concentrations. Tested specimens are described by degree of magnetic induction, which is 25 Gs and 50 Gs. Nanoparticles of both fluids are coated by oleic acid.

Measurement was realized by measurement circuit which is shown on fig. 1. As the metering instrument was used programmable electrometer Keithley 617 and the time of measurement was 1000s. Applied voltage across specimen was 100V. Acquired data was processed by Agilent HP VEE software.

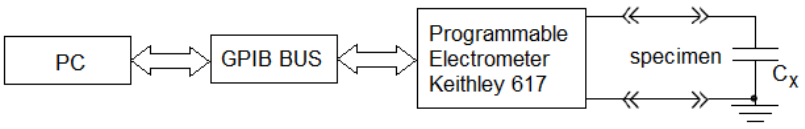


Fig. 1 Circuit configuration for measurement

The first measured sample of ferrofluid was sample with magnetic induction 25 Gs. Influences of acting external magnetic fields can be observed on these characteristics. Without this field is magnitude of current lowest at start of i(t) curves. This phenomenon is probably caused by chaotic

arrangement of nanoparticles, which are dispersed in liquid volume by interacting solvent and Brownian motion. After application of external magnetic field nanoparticles start creating long chains in direction of acting magnetic field. These chains can be long approx. 100-300 μm [2]. These chains can together create a strong structure that can have an important influenceon capacitance of specimen and therefore also charging current.

Next measured magnetic fluid was built on same carrier fluid but with double content of Fe_3O_4 particles. It caused that current i(t) is increased.

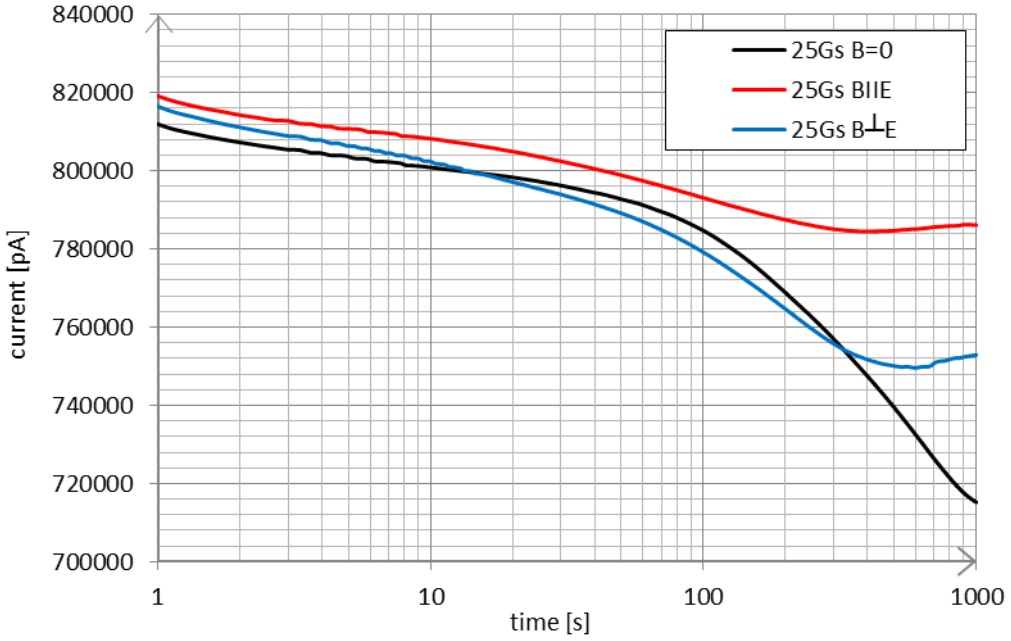


Fig. 2 i(t) characteristics of ferrofluid 25Gs

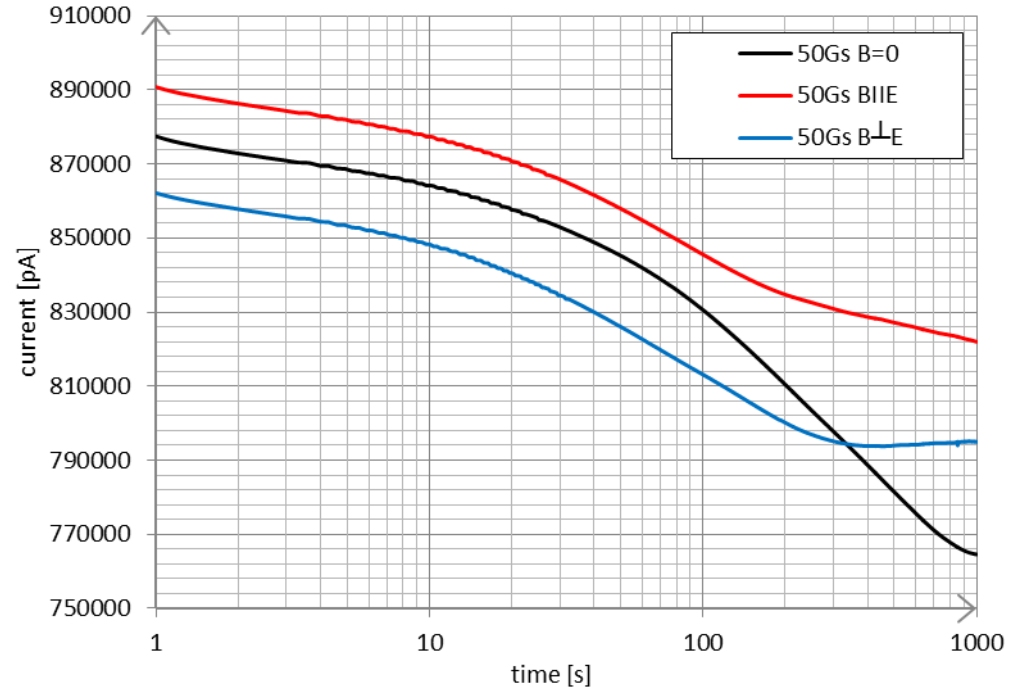


Fig. 3 i(t) characteristics of ferrofluid 50Gs

V. Conclusion

Aim of this paper was to investigate influence of arrangement of acting external magnetic field to anisotropy of i(t) characteristics of tested magnetic fluids. For confirmation of results were used two different magnetic fluids with magnetic induction 25 Gs and 50 Gs. As is showed on figures 2 and 3, concentration affected only magnitude of measured current, but not the trend of these curves. It means that nanoparticles behavior is same in every tested fluid.

We have to consider that measured current - time dependencies are macroscopic view of ongoing processes. Aim of this paper was to verify influence of magnetic nanoparticles concentration in a liquid volume of the ongoing processes on polarization. One reason is that a larger amount of magnetite nanoparticles in the magnetic fluid beneficial effect on the formation of long chains, which affect the polarization processes in the dielectric. Magnetic anisotropy which is shown on fig. 2 - 3 confirms the assumption, that action of external magnetic field in direction of electric field causes polarization current change. It's necessary to be mentioned that dependencies aren't affected only by the action of an external magnetic field but they are affected also with direction of these field. This phenomenon can be observed in every showed dependence.

Acknowledgment

We support research activities in Slovakia / Project is cofinanced from EU funds. This paper was developed within the Project "Centrum excelentnosti integrovaného výskumu a využitia progresívnych materiálov a technológií v oblasti automobilovej elektroniky", ITMS 26220120055.

This work was supported by scientific Grant Agency of the ministry of Education of the Slovak Republic project VEGA No. 1/0487/12.



References

[1] Kopčanský, Peter et al.: Magnetic nanoparticles in magnetic fluids. In: Acta lectrotechnica et Informatica. vol.10,2010, no.3, p. 10 – 13.

[2] Timko, Milan et al.: Magnetic fluid as cooling and insulation medium for high power transformers. In: Selected Topics in Energy, Environment, Sustainable Development and Landscaping, Timisoara, Politehnica University of Timisoara, Romania, 2010. p. 321-326. ISBN 978-960-474-237-0

[3] Marton, Karol et al: Situation analysis of electric field on boundary-line of solid state electrode and gas. In: Prezgląd Elektrotechniczny. Vol. 86, no. 8, p. 307-311. ISSN 0033-2097

[4] Kolcunová, Iraida – Kurimský, Juraj – Kvakovský, Milan: Modelovanie vplyvu miesta uzemnenia vodivej ochrany statorovej cievky na rozvoj povrchových výbojov. In: Elektroenergetika 2009, Košice: TU, 2009. p. 287-290. ISBN 9788055302379.

[5] Kurimský, Juraj – Dolník, Bystrík, Recovery voltange in ZnO ceramics. In: Prezgląd Elektrotechniczny. Vol. 87, no. 6, p. 214-217. ISSN 0033-2097

Skúšky digitálnych ochrán prístrojom OMICRON CMC

Miroslav KMEC, Marek HVIZDOŠ

Katedra elektroenergetiky, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

miroslav.kmec@student.tuke.sk, marek.hvizdos@tuke.sk

***Abstrakt* — Tento príspevok sa zaoberá skúšaním digitálnych ochrán z viacerých aspektov. Popisuje problematiku skúšok ochrán z bezpečnostného a prevádzkového hľadiska. V príspevku je uvedený opis testovacieho zariadenia OMICRON CMC a jeho obslužného softvéru Test Universe. Predložený príspevok poukazuje na spôsoby praktického využitia testovacieho zariadenia. Dosiahnuté výsledky svedčia o kvalite spracovania nameraných údajov a vhodnosti využitia tohto zariadenia pri testovaní terajších zložitých digitálnych ochranných systémov.**

***Kľúčové slová* — digitálne ochrany, skúšky ochrán, OMICRON CMC, Test Universe**

I. Úvod

Ochrany hrajú kľúčovú úlohu v prevádzkovaní dnešných elektrizačných sústav z hľadiska ich spoľahlivosti a bezpečnosti. Týka sa to všetkých elektrických sietí od prenosových, distribučných až po rozvody v premyslených podnikoch. Pre zabezpečenie správneho fungovania týchto sietí je potrebné použiť vhodný typ ochrany. Poruchové stavy, rovnako ako stabilné prevádzkové podmienky a spínanie je potrebné simulovať, aby sa zabezpečilo, že systém ochrán pracuje spoľahlivo ešte pred uvedením do prevádzky. Tieto skúšky sa musia po určitej dobe opakovať vzhľadom na nečinnosť ochranných zariadení za normálnych prevádzkových podmienok. Opakovaním týchto testov v určitých intervaloch sa zistí, či ochrana pracuje správne počas svojej životnosti.

V súčasnosti existuje veľa problémov v oblasti testovania ochrán. Moderné multifunkčné ochrany majú vysoký počet integrovaných funkcií. Preto sa stávajú veľmi zložitými zariadeniami. Okrem toho, rôzni dodávatelia systémov ochrán majú rôzne prevádzkové algoritmy. Tieto problémy riešia dnešné moderné testovacie súpravy, pomocou ktorých sa vykonávajú skúšky pre všetky generácie ochrán. Jedným z takýchto zariadení je aj OMICRON CMC.

II. Testovanie digitálnych ochrán a uvedenie do prevádzky

Testovanie programov ochranných zariadení predstavuje rad problémov. To preto, že hlavná funkcia ochrán sa zaoberá prevádzkou za podmienok poruchy systému a nemôže sa ľahko otestovať v bežných prevádzkových podmienkach. Túto situáciu komplikuje rastúca zložitosť systémov ochrán a obslužného softvéru ochrán. Testovanie ochranných zariadení môžeme rozdeliť do štyroch fáz [1]:

- Typové skúšky
- Rutinné testy vo výrobe
- Testy pri uvedení do prevádzky
- Pravidelné údržbové testy

Typové skúšky – tieto skúšky musia preukázať, že ochrana spĺňa zverejnené špecifikácie a je v súlade so všetkými príslušnými štandardami. Vzhľadom na skutočnosť, že hlavnou funkciou ochrany je správne pôsobenie za abnormálnych podmienok, je dôležité, aby sa výsledky posudzovali podľa týchto podmienok. Komplexné typy testov simulujúcich prevádzkové podmienky sú preto vykonávané u výrobcu počas rozvoja a certifikácie zariadení.

Typová skúška je veľmi zložitým procesom u digitálnych ochrán, ktorý sa týka testovania softvéru i hardvéru.

Rutinné testy vo výrobe – sa uskutočňujú kvôli potvrdeniu, že v priebehu výroby sa na ochrane nevyskytli žiadne nedokonalosti. Testovanie prebieha v niekoľkých fázach počas výroby, aby sa minimalizoval čas pre opravu zistených chýb.

Testy pri uvedení do prevádzky – tieto testy sú navrhnuté tak, aby dokázali, že konkrétne systémy ochrany boli správne nainštalované pred spustením do prevádzky. Všetky aspekty systému sú dôkladne preskúmané z hľadiska správnej inštalácie zariadenia, kontroly zapojenia cez kontrolu prevádzky jednotlivých položiek vybavenia. Nakoniec sa testuje celý systém.

Pravidelné údržbové testy – tieto testy sú potrebné pre identifikáciu porúch a degradáciu zariadenia v prevádzke tak, aby mohli byť prijaté nápravne opatrenia. Pretože systém ochrany funguje len pri poruchových stavoch, nedostatky nemôžu byť zistené skôr, kým nedôjde k poruche. Pravidelné testovanie pomáha pri zisťovaní nedostatkov, ktoré by boli inak odhalené až pri poruche.

A. Elektrické typové skúšky

Rôzne elektrické typové skúšky musia byť vykonané nasledovne [1]:

Funkčné testy – funkčné skúšky pozostávajú z výberu vhodných vstupov do ochrany pri teste a vyhodnotenia merania vzhľadom k predloženej špecifikácii. Zvyčajne sa vykonávajú pri kontrolovaných environmentálnych podmienkach. Testovanie môže byť rozsiahle, aj keď sa testuje len jednoduchá ochrana. Ide o časovo náročnú úlohu a na jej vykonanie je potrebné mnoho technikov. Realizácia takejto skúšky je aj finančne náročná. Pri moderných digitálnych ochranách s mnohými funkciami musí byť každá typovo skúšaná, kde funkčná skúška typového testu je hlavným problémom. Automatizované postupy sú jednoznačne potrebné pre urýchlenie takéhoto testu. Medzi ďalšie elektrické typové skúšky patria:

- Tepelná odolnosť
- Zaťaženie ochrany
- Izolačný odpor
- Pomocné napájanie

B. Skúšky elektromagnetickej kompatibility

Skladajú sa s mnohých testov, ktoré sa vykonávajú so zámerom určiť schopnosť odolnosti ochrany na elektrické prostredie, v ktorom je nainštalovaná. Prostredie rozvodni je veľmi nepriaznivé z hľadiska elektrického a elektromagnetického rušenia, ktoré tam môže vzniknúť. Existuje veľa zdrojov rušenia. Niektoré sú vnútorné, ďalšie sa prenášajú po vonkajších alebo káblových vedeniach do stanice z vonkajšieho prostredia.

Celý súbor testov je vykonávaný ako simulácia týchto druhov rušenia, ktoré spadajú do širokej škály známej pod názvom EMC, resp. testy elektromagnetickej kompatibility. Všeobecne povedané EMC je definované ako „schopnosť zariadenia koexistovať v jednom elektromagnetickom prostredí“. Elektromagnetické rušenie môže spôsobiť vážne problémy, a to je potrebné zohľadniť pri návrhu nových elektrických zariadení. Pomocou EMC testov môžeme určiť vplyvy na ochranu pri vysokofrekvenčných rušeniach rôznych druhov. Ochrana vyrábaná alebo určená pre použitie v elektrickej stanici musí byť v súlade so smernicami EÚ, určenými pre túto oblasť [1].

C. Bezpečnostné typy testov

Množstvo skúšok sa uskutočňuje kvôli potvrdeniu, že výrobok je bezpečný pri použití pre danú aplikáciu. Základné požiadavky sú, že ochrana je bezpečná a nespôsobí úraz elektrickým prúdom alebo nebezpečenstvo požiaru za normálnych podmienok a za prítomnosti jednej poruchy. K týmto skúškam patria [1]:

- Napäťová odolnosť dielektrík
- Skúšky izolácie pri prepätiach
- Impedancia ochranného pospájania

D. Environmentálne typy testov

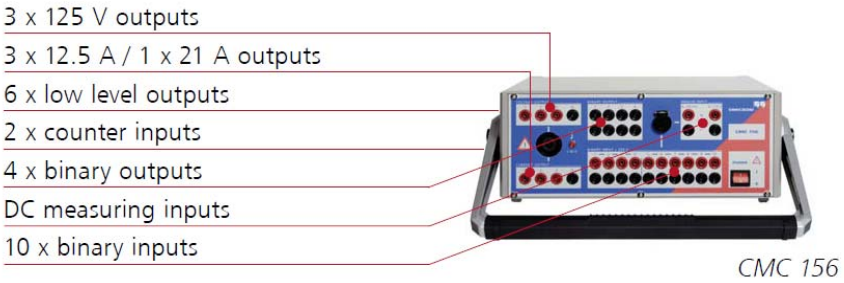
Tieto typy testov sa vykonávajú, aby sa preukázalo, že ochrana vydrží účinky životného prostredia v ktorých bude predpokladaná jej práca. Skladajú sa z nasledujúcich testov [1]:

- Teplotné testy
- Testy vlhkosti

- Testy krytím
- Mechanické testy

III. TESTOVACIE ZARIADENIE OMICRON CMC 156

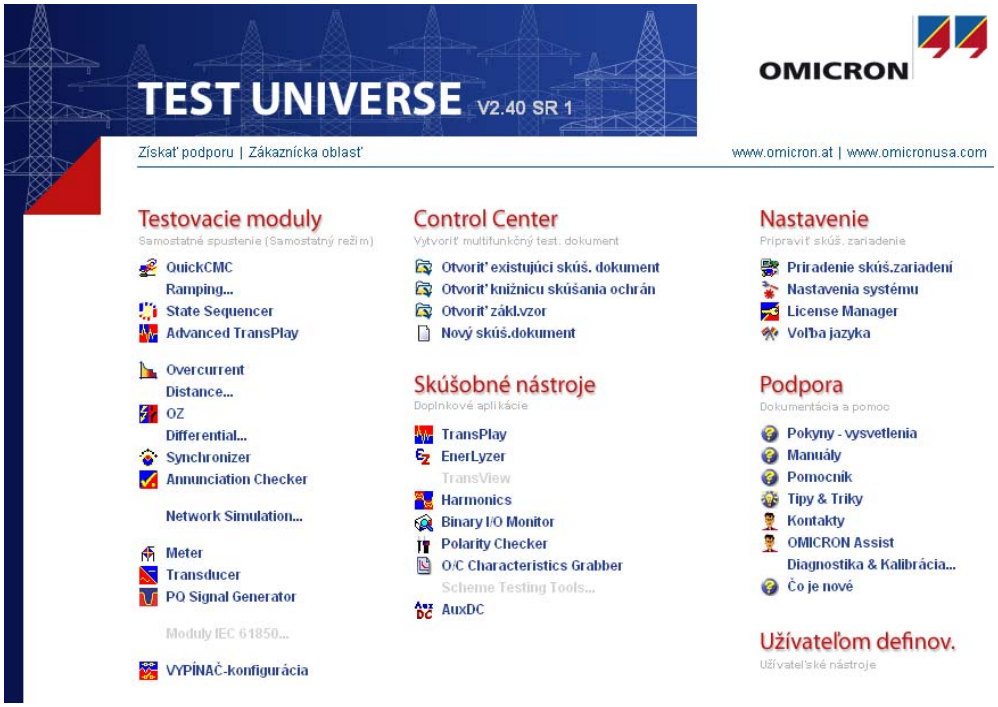
OMICRON CMC 156 (Obr. 1) je testovacie zariadenie ochrán, ktoré má potrebné vlastnosti a vynikajúcu kvalitu svojich skúšobných súprav. Oproti starším testovacím zariadeniam má lepšiu flexibilitu, presnosť, spoľahlivosť a prenosnosť. Elektronický dizajn vnútorných zosilňovačov a použitý spínaný zdroj napájania zabezpečujú minimálnu hmotnosť a objem zariadenia [2].



Obr. 1 Testovacie zariadenie OMICRON CMC 156

Skúšobné zariadenie CMC 156 je ovládané osobným počítačom. Na **PC** je spustený program **Test Universe**, ktorý komunikuje s **CMC 156** cez dodávaný **spojovací kábel**. Po nastavení komunikácie sa môže pripojiť CMC 156 k **meranému objektu** (ochrane). Využitie týchto komponentov je nevyhnutnou podmienkou správnej činnosti CMC 156.

Softvér Test Universe poskytuje plný rozsah funkcií pre definovanie a vykonanie komplexných skúšok akejkoľvek ochrany podľa smerníc výrobcu alebo skutočného nastavenia a použitia ochrany. Tento program disponuje širokou škálou testovacích modulov (Obr. 2), ktoré môžu fungovať buď samostatne, alebo môžu byť vložené do testovacích plánov pre plne automatizované testovanie. Testovacie moduly boli vytvorené pre automatizáciu skúšok. Tieto moduly poskytujú automatické vyhodnotenie a vygenerovanie skúšobného protokolu [4].



Obr. 2 Základné prostredie testovacieho softvéru Test Universe

IV. TESTOVANIE OCHRANNÝCH FUNKCIÍ DIGITÁLNEJ OCHRANY REF 543

Terminál vývodového poľa REF 543 od firmy ABB je určený na chránenie, meranie, monitorovanie a ovládanie vývodu v sieťach VN. Tento terminál disponuje nadprúdovou a podpäťovou ochrannou funkciou.



Obr. 3 Prebiehajúce meranie na ochrane REF 543

Pri skúšaní tohto terminálu boli vyhodnocované nadprúdové a podpäťové ochranné funkcie a ich vzájomné kombinácie. Nastavené parametre ochrany sú uvedené v nasledujúcich tabuľkách.

Tabuľka 1
Parametre nadprúdovej funkcie ochrany REF 543

Názov	Vypínacia charakteristika	Rozbehová hodnota	Časový index / Vypínací čas	Smer
1	IEC Extremely Inverse	0,70 Iref	0,8	Nesmerové
2	IEC Definite Time	8,20 Iref	0,15 s	Nesmerové

Tabuľka 2
Parametre podpäťovej funkcie ochrany REF 543

Názov	Vypínacia charakteristika	Rozbehová hodnota	Vypínací čas	Smer
1	IEC Definite Time	0,8 Un	3 s	-

Tabuľka 3
Parametre kombinácie podpäťovej a nadprúdovej funkcie ochrany REF 543

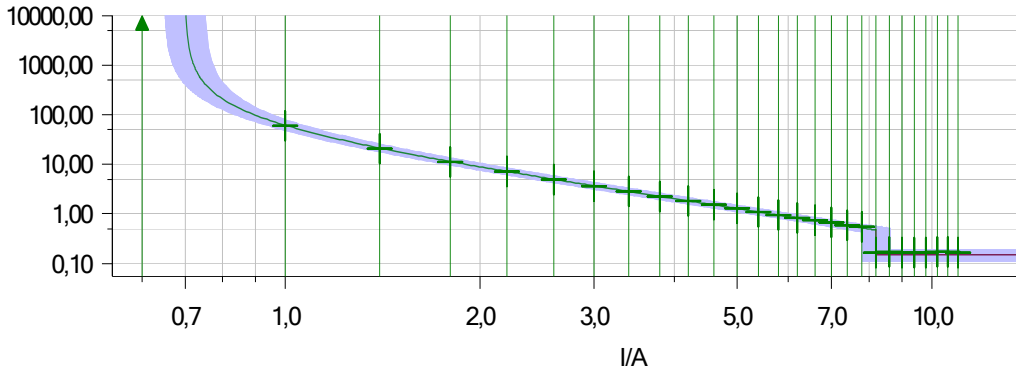
Názov	Vypínacia charakteristika	Rozbehová hodnota	Vypínací čas	Smer
1	IEC Definite Time (*U*)	0,8 Un	0,5 s	-
2	IEC Definite Time (*I*)	6,2 Iref	0,2 s	Nesmerové

Pri meraní bol na vstup ochrany privedený jeden prúdový a jeden napäťový výstup z CMC 156 a monitoroval sa popud na výkonový vypínač pomocou binárnych vstupov (kontakty bez napätia). Parametre charakteristík sa zadávali pomocou predného panela terminálu.

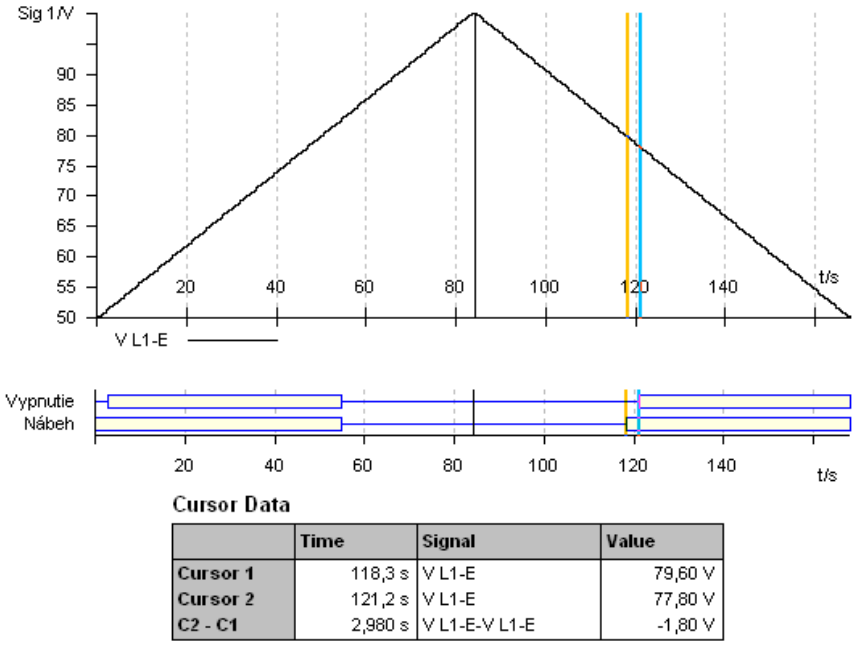
Ako prvá bola skúšaná **nadprúdová ochranná funkcia**. Na výber boli štandardné aj neštandardné typy charakteristík. Nadprúdová funkcia sa skladala z dvoch stupňov, a to $I_{>}$ a $I_{>>}$ (Extremely Inverse + Definite Time) s parametrami nastavenými podľa Tab. 1. Táto funkcia bola overovaná v Test Universe modulom Overcurrent, do ktorého sa zadali totožné parametre ako u ochrany REF. Výsledný priebeh kombinovanej dvojstupňovej vypínacej charakteristiky je na Obr. 4.

Podpäťová ochranná funkcia bola meraná modulom Ramping. Nastavila sa vstupná hodnota napätia 50 V a postupným krokom 300 mV po 0,5 s sa zvyšovala až na hranicu 100 V.

Ochrana ihneď zaznamenala podpätie a pôsobila za čas 3,025 s Po dosiahnutí hranice 83 V ochrana vypla popud na výkonový vypínač. Postupne rovnakým krokom z hranice 100 V začala klesať až na hodnotu 50 V kde pri 79,6 V zaznamenala podpätie a do času 2,98 s vyslala popud na vypnutie . Výsledná grafická závislosť je na Obr. 5.



Obr. 4 Výsledný priebeh testu nadprúdovej funkcie pomocou modulu Overcurrent



Cursor Data

	Time	Signal	Value
Cursor 1	118,3 s	V L1-E	79,60 V
Cursor 2	121,2 s	V L1-E	77,80 V
C2 - C1	2,980 s	V L1-E-V L1-E	-1,80 V

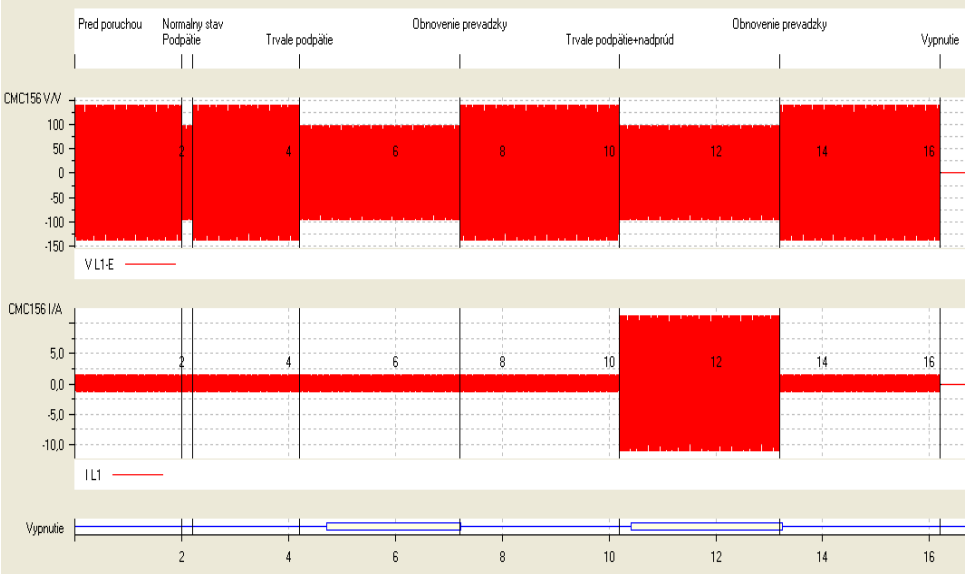
Obr. 5 Výsledný priebeh testu podpäťovej funkcie pomocou modulu Ramping

Posledný test bol zameraný na odskúšanie **kombinácie podpäťovej a nadprúdovej** ochrannej funkcie. Bol vytvorený rad sekvencií (Obr. 6) v module State sequencer, ktorými sa simulovali rôzne hodnoty vstupných veličín (U a I):

- Prvý stav (pred poruchou): I = 1 A, U = 100 V po dobu 2 s. Ochrana nepôsobí.
- Druhý stav (podpätie): I = 1 A, U = 70 V po dobu 200 ms. Podpäťová ochrana zaregistrovala podpätie ale nevypla, pretože vypínací čas ochrany bol menší ako čas trvania poruchy.
- Tretí stav (normálny stav): I = 1 A, U = 100 V po dobu 2 s. Ochrana nepôsobí.
- Štvrtý stav (trvalé podpätie): I = 1 A, U = 70 V po dobu 3 s. Podpäťová ochrana zapôsobila a vypla za čas t = 516,6 ms.
- Piaty stav (obnovenie prevádzky): I = 1 A, U = 100 V po dobu 2 s. Ochrana nepôsobí.
- Šiesty stav (trvalé podpätie + nadprúd): I = 8 A, U = 70 V po dobu 3 s. Podpäťová aj nadprúdová ochrana zaregistrovali poruchu. Vzhľadom na to, že vypínací čas

nadprúdovej ochrannej funkcie bol kratší ako u podpäťovej funkcie ochrana vypla za čas t = 216,5 ms.

- Siedmy stav (obnovenie prevádzky): I = 1 A, U = 100 V po dobu 3 s. Ochrana nepôsobí.
- Ôsmy stav (vypnutie): I = 0 A, U = 0 V.



Obr. 6 Výsledný priebeh testu kombinácie podpäťovej a nadprúdovej funkcie pomocou modulu State Sequencer

V. ZÁVER

Správna činnosť elektrických ochrán je nutnou podmienkou na dosiahnutie spoľahlivej a bezpečnej prevádzky elektrizačnej sústavy. Na testovanie digitálnych ochrán sa v elektroenergetickej praxi používajú rôzne skúšobné zariadenia. Tento príspevok poukazuje na mnohostranné využite testovacieho zariadenia OMICRON CMC a jeho obslužného softvéru Test Universe. Jeho možnosti hardvérového rozšírenia a neustále zdokonaľovanie softvérového vybavenia ho predurčujú na vykonávanie komplexných testov zložitých digitálnych ochranných systémov aj v budúcnosti.

LITERATÚRA

[1] AREVA Transmission and Distribution, Protection and Control: Relay testing and commissioning. 1987.
[2] CMC 156, Popis a obsluha. AP Consulting, 2003.
[3] Omicron Test Universe 1.4 Protection. Softwarový balík pre skúšanie ochrán. 2001.
[4] Test Universe software, PC software suite for advanced secondary testing. [Online] <http://www.omicron.at/en/products/pro/secondary-testing-calibration/sw/>.
[5] V. Chladný: *Digitálne ochrany v ES*. Košice. 2007, ISBN 978-80-8073-798-6.
[6] F. Janíček – V. Chladný – A. Beláň – Ž. Eleschová: *Digitálne ochrany v elektrizačnej sústave*. Bratislava, 2004, ISBN 80-227-2135-2.
[7] V. Chladný: *Ochrany v elektrizačných sústavách, návody na cvičenia*. Košice, 1992, ISBN 80-7099-133-X.

Viskozita prírodných esterov

Iraida KOLCUNOVÁ, Lýdia DEDINSKÁ, Viera VANČOVÁ

Katedra elektroenergetiky, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

iraida.kolcunova@tuke.sk, lydia.dedinska@tuke.sk

***Abstract* — V súčasnosti sa v elektrotechnických zariadeniach v prevažnej miere používajú ako izolanty minerálne oleje. Keďže sa v dnešnej dobe kladie dôraz na ekologickú nezávadnosť používaných materiálov je snaha nahradiť minerálne oleje rastlinnými. Preto v článku sme sa zamerali na merania viskozity prírodných esterov, ktoré hodnoty sme porovnávali s minerálnymi olejmi. Pre účely experimentu boli zvolené nové nenamáhané, tepelne namáhané a navlhnuté vzorky olejov.**

***Kľúčové slová* — dynamická viskozita, kinematická viskozita, minerálny olej, prírodný ester**

I. Úvod

V elektrotenergetike došlo k veľkému pokroku, čo sa týka vývoja biologicky odbúrateľných elektroizolačných kvapalín a rastlinných esterov. Pre využívanie alternatívnych izolačných kvapalín existujú viaceré dôvody. Jedným z dôvodov je zlá biologická odbúrateľnosť a negatívny vplyv minerálnych olejov a syntetických izolačných kvapalín na organizmy živej prírody. Ďalším dôvodom je, že minerálne oleje sú vyrábané z ropy, nakoľko ich prírodne náleziská sú vyčerpateľné je dôležité hľadať za nich vhodnú alternatívu, ktorá by bola na rastlinnej báze.

Nahradenie minerálnych olejov vyrábaných z ropy za rastlinné oleje vyrábané z ľahko obnoviteľných zdrojov prichádza do úvahy ako jedná z možností. Hľadanie adekvátnej ekologicky neškodnej náhrady za minerálne oleje bol aj dôvodom preskúmanie viskozity rastlinných olejov.

II. Viskozita Kvapalných Dielektrík

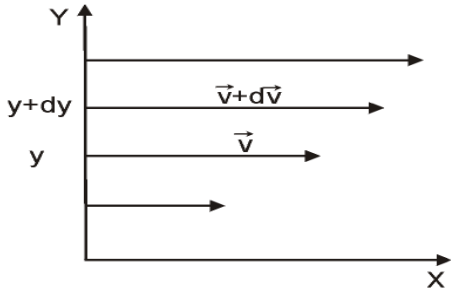
Viskozita je najdôležitejšou neelektrickou vlastnosťou kvapalných izolantov, a to ako z technologického, tak aj z funkčného hľadiska. Kvapaliny netečú rovnakou rýchlosťou ani za rovnakých podmienok. Príčinou tejto vlastnosti je vnútorné trenie (viskozita) ktoré vzniká medzi jednotlivými vrstvami prúdiacej tekutiny na rozdiel od vonkajšieho trenia, ktoré vzniká na rozhraní medzi tekutinou a stenami prostredia. Vonkajšie trenie je dosť veľké (teoreticky nekonečné, ak tekutina zmáča steny), takže vrstva tekutiny, ktorá prilieha k stene prostredia sa vôbec nepohybuje. So stúpajúcou vzdialenosťou od steny prostredia stúpa aj rýchlosť pohybu častíc kvapaliny, pričom sa pomalšie vrstvy snažia rýchlejšie vrstvy spomaľovať a opačne. Ak sa tekutina pohybuje v rúrke tak, akoby sa súosové kvapalinové valce s rovnakou rýchlosťou pohybovali za sebou, pričom vnútorne sa pohybujú rýchlejšie ako vonkajšie, potom je to laminárne prúdenie. Pri väčších rýchlostiach dochádza následkom pôsobenia dotyčnicového napätia k vzniku vírov a nepravidelnému splietaniu prúdových vlákien. Takéto prúdenie nazývame vírivé alebo tiež turbulentné [1].

Teplotnú závislosť viskozity minerálnych olejov je výhodné znázorňovať podľa Ubbelohdeho v diagrame s logaritmickou stupnicou pre teplotu a logaritmickou stupnicou pre kinematickú viskozitu v cSt, v ktorom sa zobrazí ako priamka. Priamky prislúchajúce rôznym olejom majú rôzny sklon. Na charakterizovanie strmosti teplotnej závislosti viskozity olejov sa používa tzv. viskozitný index. Na základe toho indexu sa vybrali dva oleje, jeden s veľmi strmou teplotnou závislosťou, druhý s veľmi miernou, ktoré majú rovnakú viskozitu pri 20°F (- 7°C). Viskozitami týchto olejov pri 100 °F (38°C) sú dané základné body stupnice viskozitného indexu. Určitému oleju sa potom priraďuje viskozitný index podľa toho, aký sklon má priamka vyjadrujúca jeho teplotnú závislosť pri porovnaní s priamkami oboch vzťažných olejov. Teda čím väčší viskozitný index, tým menej sa mení viskozita oleja s teplotou. Viskozitný index závisí od zastúpenia základných uhľovodíkov: naftentické a aromatické uhľovodíky viskozitný index znižujú, nerozvetvené parafinické zvyšujú [1].

A. Dynamická viskozita

Vnútorné trenie alebo viskozita kvapaliny je mierou jej vlastnosti tiecť. Kvapaliny s menšou viskozitou (voda) tečú ľahšie ako kvapaliny s väčšou viskozitou (meď). Reálne kvapaliny sa vždy vyznačujú vnútorným trením. Jeho príčinou sú väzbové sily, prostredníctvom ktorých sa k sebe viažu atómy resp. molekuly z ktorých kvapalina pozostáva. Ak tieto väzbové sily nie sú veľké, môžu sa ľahko prerušiť, takže jednotlivé čiastočky kvapaliny sa môžu po sebe posúvať a kvapalina môže tiecť. Ak sú tieto sily veľké, tečenie kvapaliny je veľmi pomalé [2].

Predstavme si kvapalinu prúdiacu v potrubí v smere osi X, (Obr. 1). Rozdeľme ju na veľmi tenké vrstvy, ktoré v závislosti od vzdialenosti od steny potrubia, t.j. od súradnice Y, sa pohybujú rôznymi rýchlosťami v, a to tak, že rýchlosť prúdenia kvapaliny sa zvyšuje s rastúcou vzdialenosťou od steny potrubia. Najvyššia je v strede potrubia, najmenšia pri jeho stenách, lebo tam sú najväčšie väzbové sily medzi časticami kvapaliny a časticami materiálu potrubia [2].



Obr. 1 Prúdenie viskozity v potrubí

Z obrázku vyplýva, že medzi jednotlivými vrstvičkami kvapaliny pohybujúcimi sa rôznou rýchlosťou vzniká tangenciálne napätie, ktoré má smer prúdenia kvapaliny. Príčinou tohto napätia je tepelný pohyb častíc, v dôsledku ktorého dochádza k prenosu častíc medzi dvoma susednými vrstvami v oboch smeroch. Častice prichádzajúce z pomalšej vrstvy do rýchlejšej vrstvy tak spomaľujú tečenie rýchlejšej vrstvy a naopak.

Na základe takéhoto modelu možno pre veľkosť tangenciálneho napätia odvodiť vzťah [2]:

$$\tau = \eta \cdot \frac{\partial v}{\partial y} \tag{1}$$

kde:

$\frac{\partial v}{\partial y}$ - gradient (zmena) rýchlostí v smere osi y,

η - koeficient dynamickej viskozity,

τ - tangenciálne napätie medzi pohybujúcimi sa vrstvičkami kvapaliny.

A. Kinematická viskozita

V technickej praxi sa používa kinematická viskozita υ, čo je pomer dynamickej viskozity η a hustoty skúšanej kvapaliny ρ pri tej istej teplote [2]:

$$v = \frac{\eta}{\rho} \tag{2}$$

Podstatou je stanovenie doby prietoku stáleho objemu kvapaliny kapilárou kalibrovaného viskozimetra pri určitej hydrostatickej výške kvapaliny a pri prísne kontrolovanej a presne meranej teplote. Kinematická viskozita sa určí zo súčinu zmeranej doby prietoku a konštanty použitého viskozimetru (najčastejšie sa používa kapilárny viskozimeter podľa Ubbel - Ondeho). Za konštantu sa dosadzuje hodnota vyznačená na viskozimetri výrobcu. V použitom oleji viskozita rastie z dôvodu vzniku nečistôt, a to buď prienikom z vonkajšieho prostredia alebo opotrebením a starnutím oleja. Pokles viskozity môže byť spôsobený zriedením oleja a vody [2].

Meraním kinematickej viskozity olejov sa sleduje či olej spĺňa svoju funkciu. Limitnou hodnotou pre zmenu kinematickej viskozity je ± 10 % (často sa udáva napr. pri meraní minerálnych olejov) [2].

III. MERANIE VISKOZITY KVAPALNÝCH DIELEKTRÍK

Na meranie viskozity rastlinných a minerálnych olejov v závislosti na teplote oleja sa použil Höpplerov viskozimeter. Merania sa vykonávali za použitia stopiek, teplomera s rozsahom od 20ºC do 100ºC, banky, pinzety a sklenenej tyčke. Zvláštna pozornosť sa musela venovať výberu guľôčky a čistote padajúcej trubice. Stopkami sa meral čas pádu guľôčky od prvej kruhovej značky meracej rúrky po druhú kruhovú značku meracej rúrky. Meranie sa opakovalo pre každú hodnotu teploty.

Dynamická viskozita sa počíta podľa vzorca [4]:

$$\eta = t \cdot (\rho_1 - \rho_2) \cdot K \tag{3}$$

Kde:

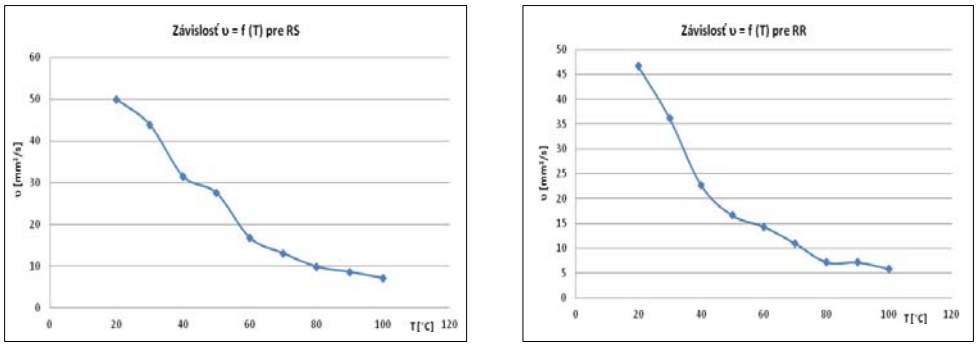
η – dynamická viskozita v milipascalsekundách [mPa.s]
t – doba padanie guľôčky v sekundách [s]
ρ_1 – hustota guľôčky [g.cm^{-3}]
ρ_2 – hustota kvapaliny pri teplote merania [g.cm^{-3}]
K – konštanta guľôčky [mPa.s.cm^3.g^{-1}.s^{-1}]

Hodnoty ρ_1 a K sú pre každú guľôčku uvedené v kontrolnom osvedčení. Hodnota (ρ_1 -ρ_2) vyžaduje korektúru tlaku. Pri používaní sklenených guľôčok sa odporúča uvažovať hustotu kvapaliny na dve jednotky desatinného miesta. U kovových guľôčok stačí presnosť na jednu jednotku druhého desatinného miesta, čo sa dá ľahko odhadnúť. Prepočet dynamickej viskozity na kinematickú viskozitu sa určil podľa vzorca (2).

Meranie viskozity sa vykonalo na nasledujúcich izolačných kvapalinách:

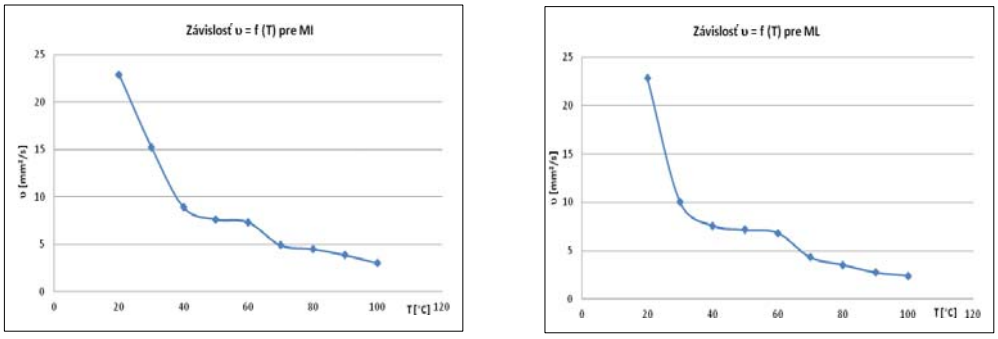
- rastlinný slnečnicový olej filtrovaný a chemicky upravovaný (rafinovaný) pre použitie v potravinárskom priemysle: (**RS**)
- rastlinný repkový olej filtrovaný a chemicky upravovaný (rafinovaný) pre použitie v potravinárskom priemysle: (**RR**)
- inhibovaný transformátorový olej ITO 100 vysoko rafinovaný minerálny olej s prísadou 2,6-di-terc-butyl-metylfenolu: (**MI**)
- inhibovaný transformátorový olej Nynas – Lyra X: (**ML**)

Výsledky merania jednotlivých typov olejov v závislosti na teplote pre rastlinné oleje sú uvedené na obrázku (Obr. 2)a pre minerálne oleje na (Obr. 3).



a.) RS b.) RR

Obr. 2 Závislosť viskozity υ od teploty T pre a.) rastlinný slnečnicový olej RS, b.) rastlinný repkový olej RR



a.) MI b.) ML

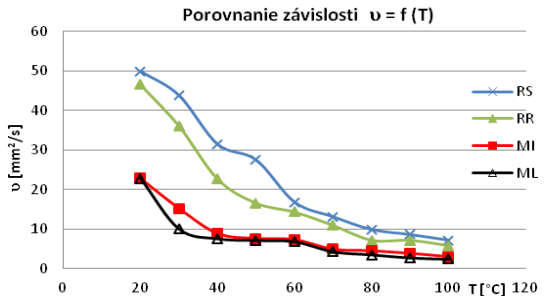
Obr. 3 Závislosť viskozity υ od teploty T pre a.) minerálny olej ITO 100, b.) minerálny olej LYRA X

Z uvedených závislostí je možné konštatovať, že pre všetky merané vzorky oleja hodnota viskozity zo zvyšujúcou sa teplotou klesá. Pri zmene teploty od 20°C do 40°C pokles viskozity je výraznejší ako pri ďalšom zvyšovaní teploty. Hodnoty viskozity pre vzorky olejov v rozsahu teplôt od 20 °C do 100 °C sú uvedené v tabuľke (Tab. 1). V tepelnom rozsahu od 20°C do 100°C klesne kinematická viskozita pre:

- rastlinný slnečnicový olej RS o 43 $mm^2.s^{-1}$
- rastlinný repkový olej RR o 39 $mm^2.s^{-1}$
- minerálny olej ITO 100 MI o 19 $mm^2.s^{-1}$
- minerálny olej Lyra X ML o 20 $mm^2.s^{-1}$

Tabuľka 1
Hodnoty viskozity υ kvapalných dielektrík pre rôzne teploty T

T [°C]	υ [mm^2/s]			
	RS	**RR**	**MI**	**ML**
20	49,92167	46,65252	22,88493	22,81466
30	43,78997	36,14536	15,25873	10,04429
40	31,45965	22,72854	8,905008	7,580025
50	27,4975	16,60845	7,640402	7,147243
60	16,79743	14,34059	7,33852	6,836229
70	13,1041	10,96654	4,916362	4,305495
80	9,865464	7,197823	4,509709	3,528213
90	8,598768	7,182093	3,876011	2,744842
100	7,152215	5,827184	3,022878	2,416323



Obr. 4 Závislosť viskozity υ od teploty T pre kvapalné dielektriká

Pri porovnaní viskozity kvapalných dielektrík (Obr. 4) je vidieť že pri nižších teplotách rozdiel v hodnote viskozity medzi rastlinnými a minerálnymi olejmi je približne 25 $mm^2.s^{-1}$. Pri zvyšovaní teploty dochádza k zmenšovaniu rozdielu hodnôt viskozít minerálnych a rastlinných olejov a pri teplote 100°C rozdiel činí len 3 $mm^2.s^{-1}$.

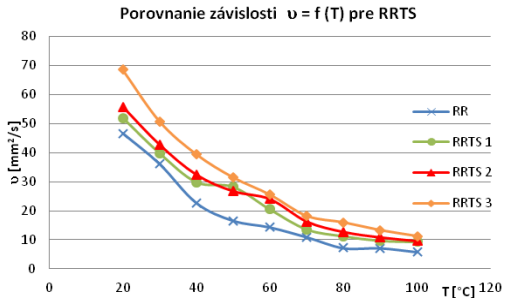
IV. Meranie Viskozity Tepelne Zostarnutého Oleja

Pre určenie zmien viskozity v dôsledku tepelného namáhania boli zvolené vzorky rastlinného repkového oleja (RR) a minerálneho Lyra X (ML). Väčšie množstvo oleja bolo umiestnené do teplovzdušnej sušičky, kde sa dlhodobo namáhalo pri teplote T = 80°C. V určitých časových intervaloch bola časť oleja odobratá na meranie, zvyšná časť sa namáhala ďalej.

Pre lepší prehľad nameraných výsledkov merané vzorky boli označené nasledovne:

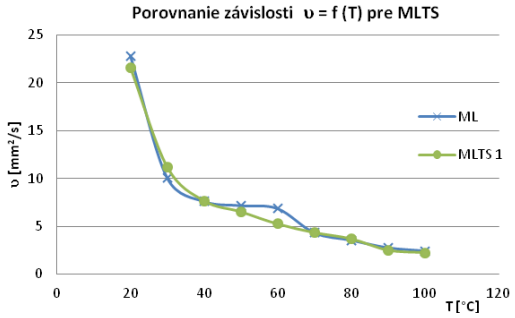
- rastlinný repkový olej s časom namáhania t = 850 h (**RRTS 1**)
- rastlinný repkový olej s časom namáhania t = 5 000 h (**RRTS 2**)
- rastlinný repkový olej s časom namáhania t = 8 500 h (**RRTS 3**)
- minerálny olej LYRA X s časom namáhania t = 850 h (**MLTS 1**)

Teplotná závislosť viskozity pre tepelne zostarnutý rastlinný repkový olej RR je znázornená na obrázku (Obr. 5) a pre minerálny olej Lyra X ML na obrázku (Obr.6). Pre všetky vzorky olejov viskozita so zvyšujúcou sa teplotou klesá. Pri nižších teplotách pokles hodnoty viskozity je výraznejšie ako pri vyšších teplotách, kde pokles viskozity sa pomaly zmenšuje.



Obr. 5 Závislosť viskozity υ od teploty T pre tepelne zostarnutý rastlinný repkový olej RRTS

Hodnota viskozity tepelne zostarnutých vzoriek rastlinných olejov je vyššia v porovnaní s nenamáhanou vzorkou oleja. S narastajúcou dobou namáhania hodnota viskozity rastie.



Obr. 6 Závislosť viskozity υ od teploty T pre tepelne zostarnutý minerálny olej ML

Priebeh viskozity namáhaného MLTS 1 a nenamáhaného ML minerálneho oleja sú identické. Tepelne namáhanie pri teplote 80 °C po dobu 850 hodín neovplyvnilo hodnotu viskozity minerálneho oleja Lyra X.

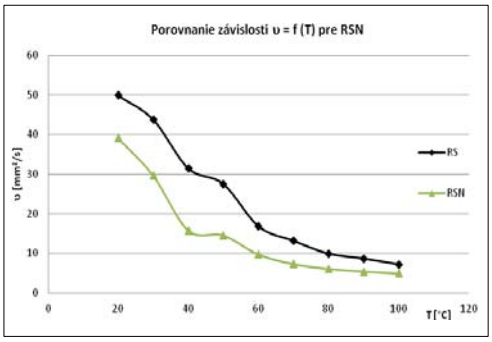
V. Meranie Viskozity Nalhnutého Oleja

Za účelom zistenia zmeny viskozity v dôsledku navlhnutia boli vybraté vzorky olejov vystavené atmosférickým podmienkam po dobu trvania 2 160 hodín. Na skúšku sa vybrala vzorka rastlinného slnečnicového oleja (RS), rastlinného repkového oleja (RR), minerálnych olejov ITO 100 (MI) a Lyra X (ML). Po 2 160 hodinách vystavenia vo voľne uzatvorených nádobách vonku boli na vzorkách oleja uskutočnené merania a výsledky meraní sa porovnali s pôvodnými vzorkami.

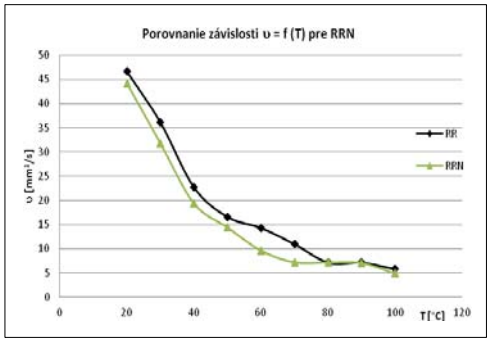
Označenie vzoriek olejov, ktoré boli vystavené atmosférickým podmienkam:

- navlhnutý rastlinný slnečnicový olej: (**RSN**)
- navlhnutý rastlinný repkový olej: (**RRN**)
- navlhnutý minerálny olej ITO 100: (**MIN**)
- navlhnutý minerálny olej Lyra X: (**MLN)**

Porovnanie teplotnej závislosti viskozity navlhnutej a nenavlhnutej vzorky rastlinných a minerálnych olej je znázornené na obrázku (Obr. 7 a Obr.8).

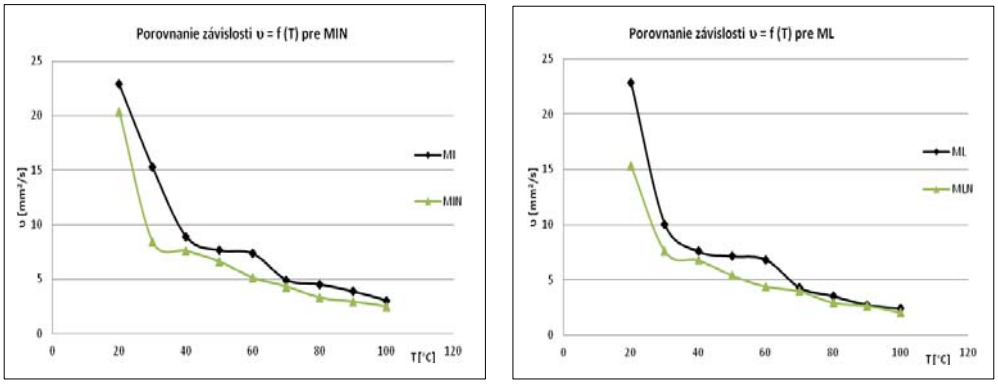


a.) RSN



b.) RRN

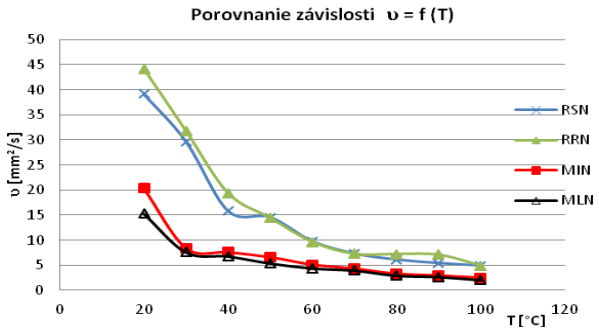
Obr. 7 Závislosť viskozity υ od teploty T pre a.) navlhnutý rastlinný slnečnicový olej RSN, b.) navlhnutý rastlinný repkový olej RRN



a.) MIN b.) MLN

Obr. 8 Závislosť viskozity od teploty T pre a.) navlhnutý minerálny olej ITO 100, b.) navlhnutý minerálny olej Lyra X

Pre všetky vzorky olejov je možne konštatovať, že vlhkosť znížila viskozitu olejov. So zvyšujúcou sa teplotou dochádza k odparovaniu vody a rozdiel v hodnotách viskozít sa zmenšuje. V prípade rastlinného repkového oleja a minerálneho oleja LYRA X od 80°C hodnoty viskozít navlhnutej a nenavlhnutej vzorky sa vyrovnali. Najväčšie rozdiely sú zaznamenané u rastlinného slnečnicového oleja, kde viskozita navlhnutej vzorky je nižšia v celom teplotnom rozsahu.



Obr. 9 Závislosť viskozity υ od teploty T pre všetky vzorky navlhnutých olejov

Z porovnania teplotnej závislosti viskozity navlhnutých vzoriek olejov (Obr. 9) je vidieť že priebeh zmeny viskozity sa podoba na závislosti nenavlhnutých vzoriek olejov (Obr. 4).

VI. ZÁVER

Hodnota viskozity rastlinných olejov je vyššia ako minerálnych. S narastajúcou teplotou viskozita všetkých meraných vzoriek olejov klesá. Pri izbovej teplote rozdiel v hodnote viskozít rastlinných a minerálnych olejov je veľký a sa so stúpajúcou teplotou zmenšuje. Nad teplotou 70 °C hodnota viskozity rastlinných a minerálnych olejov je porovnateľná. Tepelné namáhanie pri teplote 80 °C spôsobilo zvýšenie viskozity rastlinného oleja a neovplyvnilo hodnotu viskozity minerálneho oleja. Vplyvom navlhnutia hodnota viskozity klesla v porovnaní s nenavlhnutou vzorkou ako u rastlinných tak aj u minerálnych olejov.

POĎAKOVANIE

Tento článok bol vypracovaný s podporou projektu Vývoj unikátneho nízkoenergetického statického zdroja pre elektrosystémy, ITMS 26220220029, ktorý je spolufinancovaný zo štrukturálneho fondu EÚ ERDF v rámci výzvy OPV a V-2008/2.2/01-SORO a prioritnej osi 2 Podpora výskumu a vývoja.

LITERATÚRA

[1] Artbauer, Šedovič, Adamec: Izolanty a izolácie, ALFA, Bratislava, 1969
[2] VANČOVÁ: Vplyv teploty na viskozitu rastlinných a minerálnych olejov, Diplomová práca, Technická univerzita v Košiciach, 2011, Košice
[3] Viskozita. [cit 2012-02-10]. Dostupné na internete:: <http://sk.wikipedia.org/wiki/Viskozita>
[4] Príručka na meranie s Höpplerovým viskozimetrom : TEKO, Košice

Štatistická analýza čiastkových výbjov

Iraida KOLCUNOVÁ, Marián HRINKO, Juraj KURIMSKÝ

Katedra elektroenergetiky, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

iraida.kolcunova@tuke.sk, marian.hrinko@tuke.sk, juraj.kurimsky@tuke.sk

***Abstrakt* — Článok pojednáva o štatistickej analýze čiastkových výbojov. Z nameraných a zaznamenaných údajov sú zostrojené histogramy fázových rozložení maximálnych hodnôt zdanlivých nábojov, stredných hodnôt zdanlivých nábojov a početnosti výbojov. Z histogramov sú prostredníctvom programu vypočítané deskriptívne charakteristiky, ktoré výrazne redukujú množinu nameraných dát a umožňujú tak rýchlejšie vyhodnotenie výbojovej činnosti umelou inteligenciou. Naviac v článku sa sleduje závislosť šikmosti a špicatosti fázových rozložení od priloženého napätia.**

***Keywords* — čiastkové výboje, popisné charakteristiky, šikmosť, špicatosť, korelačný faktor**

I. Úvod

Čiastkové výboje poškodzujú izolačný systém vysokonapäťových strojov točivých, čím znižujú spoľahlivosť prevádzky stroja a zvyšujú pravdepodobnosť jeho zlyhania. Z toho dôvodu detekcia čiastkových výbojov je veľmi dôležitým nástrojom v diagnostike vysokonapäťových točivých strojoch. Patrí medzi spoľahlivé diagnostické metódy, informuje o integrite izolačného systému a nedostatkoch, ktoré vznikli počas výrobného procesu. Obrazce (vzory) čiastkových výbojov umožňujú odhaliť zdroj a pôvod ich vzniku, preto je meranie čiastkových výbojov aj pomocným nástrojom na určovanie kvality výrobného procesu izolačných systémov.

Mnoho rokov dešifrovanie vzorov čiastkových výbojov bolo vykonávané prostredníctvom expertov. Avšak pokrok v oblasti výpočtovej technike a techník rozoznávania vzorov čiastkových výbojov umožnil vytvoriť autonómny systém v rozpoznávaní vzorov čiastkových výbojov – rôzne expertné systémy na báze neurónovej siete.

Dôležitým medzikrokom k vytvoreniu autonómneho systému rozpoznávania vzorov čiastkových výbojov je štatistické spracovanie nameraných údajov, ktoré sú vstupnými informáciami do expertného systému. Namerané údaje čiastkových výbojov predstavujú veľkú množinu dát a spracovanie takého kvanta údajov by zabralo mnoho času klasifikácie vzorov čiastkových výbojov. Štatistickou analýzou sú dáta redukované do podoby číselných charakteristík, ktoré výrazne znižujú množinu vstupných údajov a ktoré v sebe zahŕňajú potrebné informácie o tvare vzoroch čiastkových výbojov.

Tento článok je zameraný na štatistické spracovanie nameraných údajov do podoby deskriptívnych charakteristík, ktoré výrazne redukujú množinu vstupných údajov a umožňujú tak rýchlejšiu klasifikáciu vzorov čiastkových výbojov.

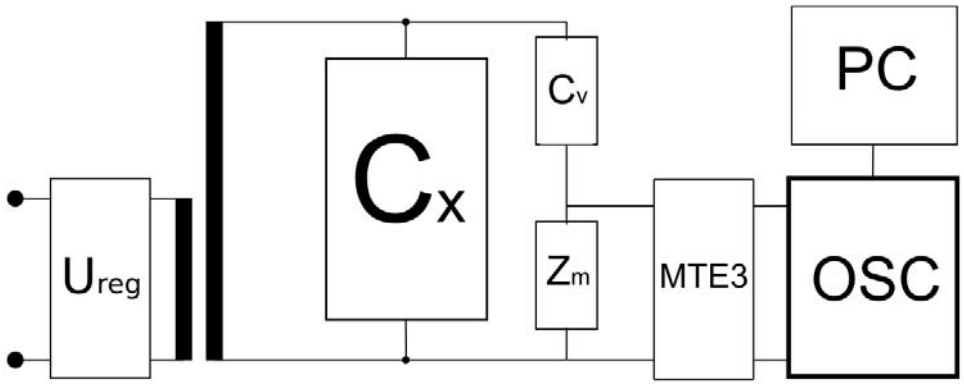
II. Praktická Časť

Meraným objektom bola vysokonapäťová cievka, ktorej menovité napätie bolo U_n = 6 kV a menovitý výkon P_n = 0,2 MW. Hlavná izolácia meranej cievky bola vytvorená z kalcinovo-sľudovej pásky spájaná epoxidovou živicou technológiou Resin Rich. Za účelom získania potrebných informácii (hodnôt zdanlivého náboja a početnosti v danom časovom okamihu) pre štatistické spracovanie bola cievka odmeraná päť krát pri rovnakých napäťových hladinách.

Meranie bolo realizované v laboratórnych podmienkach, t.j. meraný objekt - vn cievka bola zavesená na izolačnom lane vo Faradayovej klietke. Konce cievky boli spojené homogenizačnou elektródou, za účelom potlačenia korónových výbojov na ostrých výčnelkoch. Drážková časť cievky bola uzemnená prostredníctvom štyroch bodov spojených vodivou sieťovinou. Týmto spôsobom sa namodelovalo uloženie cievky v drážke statora. Bloková schéma zapojenia je na Obr. 1.

Regulačným zdrojom napätia U_{reg} sa privádzalo napätie na homogenizačnú elektródu meranej cievky C_x. Paralelne k meranému objektu bol zapojený väzobný kondenzátor C_v v sérii

s meracou impedanciou Z_m, na ktorej pri prúdových impulzoch vyvolaných čiastkovými výbojmi vznikali úbytky napätí. Tie boli zosilnené v meracom a zosilňovacom zariadení MTE3 a privádzané do digitálneho osciloskopu OSC. Osciloskopom sa analógový signál digitalizoval a ďalej posielal do personálneho počítača PC, kde sa vyhodnocovala a zaznamenávala výbojová činnosť.



Obr. 1 Bloková schema zaponenia merania čiastkových výbjov.

III. TEORETICKÝ ROZBOR

Na popis tvaru štatistického súboru nameraných údajov zdanlivých nábojov a početností boli použité charakteristiky polohy, charakteristiky variability, charakteristiky šikmosti a charakteristiky špicatosti. Okrem týchto charakteristík na posúdenie vzájomnej závislosti medzi kladnou a zápornou polperiódou napätia bola použitá charakteristika popisujúca závislosť medzi veličinami – korelačný faktor.

Prvým krokom štatistickej analýzy vzorov čiastkových výbojov bolo zostrojenie histogramov fázových rozložení maximálnych hodnôt zdanlivých nábojov H_{qmax}, stredných hodnôt zdanlivých nábojov H_{qn} a početnosti výbojov H_n. Namerané hodnoty čiastkových výbojov boli zaznamenané v dvoch súboroch. V prvom súbore boli zaznamenané amplitúdy zdanlivých nábojov, v druhom im prislúchajúca početnosť. V podstate sa jedná o maticu 440 riadkov a 199 stĺpcov, kde riadky predstavujú počet periód aplikovaného napätia a stĺpce počet fázových okienok. Z toho vyplýva, že jednému fázovému okienku prislúcha 1,81°.

Histogramy fázového rozloženia maximálnych hodnôt zdanlivých nábojov H_{qmax} zachytených v jednotlivých fázových okienkach boli vytvorené podľa:

$$H_{q\max,i} = \max_{1\le j\le n}\left\lfloor q_{i,j}\right\rfloor \tag{1}$$

Na vytvorenie histogramov fázových rozložení stredných hodnôt náboja H_{qn} a početnosti H_n výbojov bola použitá charakteristika polohy - stredná hodnota, pomocou ktorej bola vypočítaná stredná hodnota zdanlivého náboja a početnosť výbojov. Pre histogramy fázových rozložení stredných hodnôt zdanlivých nábojov platí:

$$H_{qn,i} = \frac{1}{n_i}\sum_{j=1}^{n} q_{i,j} n_{i,j} \tag{2}$$

a pre histogramy početnosti výbojov platí:

$$H_{n,i} = \frac{1}{n}\sum_{j=1}^{n} n_{i,j} = \frac{n_i}{n} \tag{3}$$

Index i označuje fázové okienko, index j označuje príslušnú hodnotu zdanlivého náboja v i-tom fázovom okienku, n je počet zachytených periód aplikovaného napätia a n_i označuje celkovú početnosť výbojov v príslušnom fázovom okienku, pričom $\sum n_{i,j}=n_i$. Keďže štatistický súbor je matica s rozmerom 440 riadkov a 199 stĺpcov, index i môže nadobúdať hodnotu 1 - 199, a index j hodnotu 1 - 440.

Čiastkové výboje sa počas napäťového cyklu vyskytujú v dvoch polperiódach, v kladnej a v zápornej polperióde priloženého napätia. Histogramy rozdelení H_{qmax}, H_{qn}, H_n sú preto charakterizované pre kladnú aj zápornú polvlnu aplikovaného napätia H^+_{qmax}, H^+_{qn},H^+_n a H^-_{qmax}, H^-_{qn}, H^-_n. Na Obr. 2 sú zobrazené histogramy fázových rozložení H_{qmax}, H_{qn} a H_n pre napäťovú hladinu 6kV.



Obr. 2 Histogramy fázových rozložení H_{qmax}, H_{qn} a H_n pri napätí 6kV.

Histogramy jednotlivých fázových rozložení zdanlivých nábojov H_{qmax}, H_{qn} a H_n predstavujú tzv. vzory čiastkových výbojov. Na zjednodušený popis histogramov H^+_{qmax}, H^+_{qn},H^+_n a H^-_{qmax}, H^-_{qn}, H^-_n boli použité číselné charakteristiky šikmosti, špicatosti, nábojovej asymetrie a korelačného faktora.

Pomocou smerodajnej odchýlky a rozptylu boli počítané charakteristiky šikmosti a špicatosti. Rozptyl (disperzia) je druhý centrálny moment definovaný vzťahom:

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}\left(q_i - \bar{q}\right)^2 n_i \tag{4}$$

pričom smerodajná odchýlka je jeho druhou odmocninou.

Operátor šikmosti je definovaný pomocou tretieho centrálneho momentu a tretej mocniny smerodajnej odchýlky podľa vzťahu:

$$\gamma_1 = \frac{m_3}{\left(\sqrt{m_2}^3\right)} = \frac{1}{n}\frac{\sum_{i=1}^{n}\left(q_i - \bar{q}\right)^3 n_i}{s^3} \tag{5}$$

Operátor šikmosti je charakteristikou rozdelenia, ktorý popisuje jeho nesymetriu. $\gamma_1 = 0$ znamená rovnomerné rozdelenie vpravo aj vľavo, $\gamma_1 > 0$ znamená naklonenie rozdelenia do ľavej strany, $\gamma_1 < 0$ znamená naklonenie rozdelenia do pravej strany.

Operátor špicatosti je definovaný pomocou štvrtého centrálneho momentu a druhej mocniny rozptylu. Kedže normálne rozdelenie má špicatosť rovnú trom, je nutné túto trojku odpočítať

$$\gamma_2 = \frac{m_4}{\left(\sqrt{m_2}^4\right)} - 3 = \frac{1}{n}\frac{\sum_{i=1}^{n}\left(q_i - \bar{q}\right)^4 n_i}{s^4} - 3 \tag{6}$$

Operátor špicatosti je charakteristikou rozdelenia, ktorý porovnáva dané rozdelenie s normálnym rozdelením. $\gamma_2 = 0$ znamená, že dané rozdelenie pravdepodobnosti je zhodné s normálnym rozdelením, $\gamma_2 > 0$ znamená , že charakteristika daného rozdelenia je ostrejšia, $\gamma_2 < 0$ znamená, že charakteristika daného rozdelenia je plochejšia.

Ďalším popisným operátorom je nábojová asymetria Q, ktorá je definovaná:

$$Q = \frac{Q_S^- / N^-}{Q_S^+ / N^+} = \frac{N^+ \sum_{i=n/2}^{n} q_i}{N^- \sum_{i=1}^{n/2} q_i} \tag{7}$$

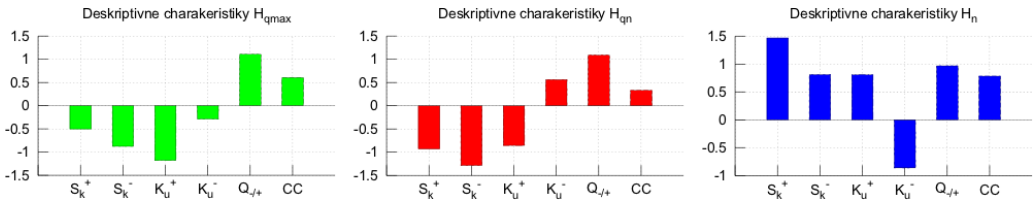
kde Q^-, Q^+ sú sumačné náboje kladnej a zápornej polperiódy a N^-, N^+ sú početnosti v kladnej a zápornej polperióde. Nábojová asymetria popisuje symetriu rozdelení medzi kladnou a zápornou polperiódou.

Pri výpočte korelačného koeficientu bolo potrebné vypočítať kovariáncie pre kladnú polperiódu, zápornú polperiódu a vzájomnú kovarianciu kladnej a zápornej polperiódy. Výsledný vzťah na výpočet korelačného faktora je:

$$cc = \frac{\sum_{n}^{n/2} q_{i+n/2} q_i - n\left(\bar{q}\right)^-\left(\bar{q}\right)^+}{\sqrt{\left(\sum_{i=n/2}^{n} q_i^2 - n\left[\left(\bar{q}\right)^-\right]^2\right)\left(\sum_{i=1}^{n/2} q_i^2 - n\left[\left(\bar{q}\right)^+\right]^2\right)}} \tag{8}$$

kde n je počet stĺpcov štatistického súboru, t.j. 199, $(\bar{q})^-$ a $(\bar{q})^+$ sú stredné hodnoty nábojov zápornej a kladnej polperiódy. Korelačný faktor popisuje koreláciu medzi kladnou a zápornou polperiódou rozdelení. cc = 0 znamená, že medzi kladnou a zápornou polperiódy rozdelenia nie je lineárna závislosť, cc = -1 znamená, že medzi kladnou a zápornou polperiódou rozdelenia je záporná lineárna závislosť, cc = +1 znamená, že medzi kladnou a zápornou polperiódou rozdelenia je kladná lineárna závislosť.

Na Obr. 3 sú zobrazené grafy vypočítaných deskriptívnych charakteristík histogramov fázových rozložení H^+_{qmax}, H^+_{qn},H^+_n a H^-_{qmax}, H^-_{qn}, H^-_n pri 6kV napätí.



Obr. 3 Popisné charakteristiky histogramov fázových rozložení H_{qmax}, H_{qn} a H_n pri napätí 6kV

Popisné charakteristiky histogramov fázových rozložení maximálnych hodnôt, stredných hodnôt a početnosti výbojov pre napätie 6kV.

IV. Zhodnotenie Merania

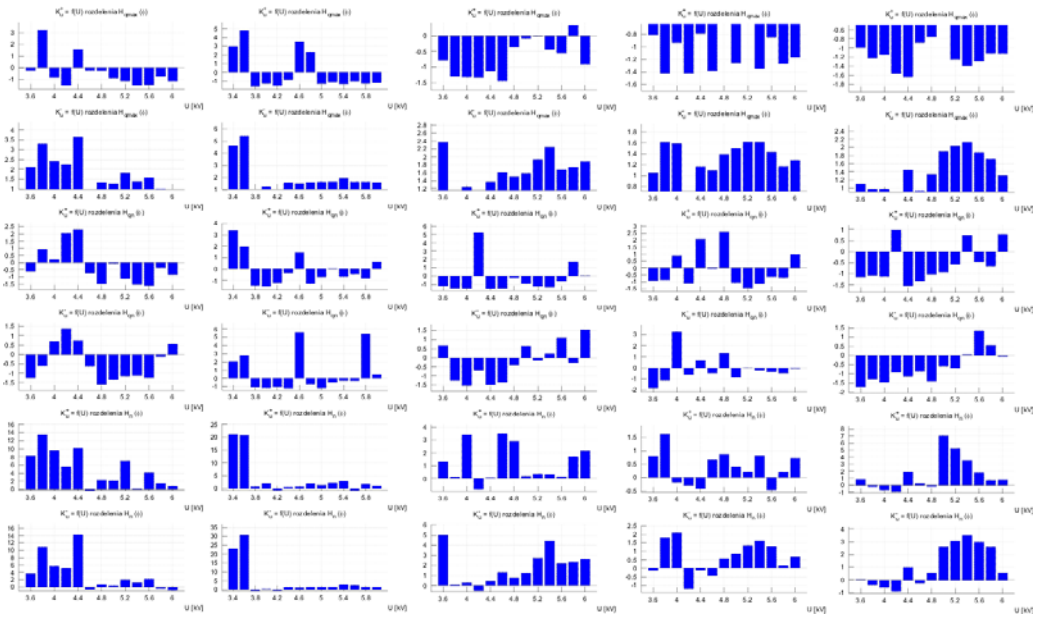
Okrem štatistického spracovania nameraných údajov, bola v tomto článku sledovaná aj závislosti šikmosti a špicatosti histogramov fázového rozloženia H^+_{qmax}, H^+_{qn},H^+_n a H^-_{qmax}, H^-_{qn}, H^-_n od priloženého napätia. Na Obr. 4 a 5 sú zobrazené závislosti šikmosti a špicatosti histogramov H^+_{qmax}, H^+_{qn},H^+_n a H^-_{qmax}, H^-_{qn}, H^-_n od priloženého napätia pre všetkých päť meraní.



Obr. 4 Hodnoty šikmosti vykonaných meraní v závislosti od napätia

Ako vidieť na Obr. 4 hodnoty šikmosti histogramov rozdelení H_{qmax}, H_{qn} a H_n so vzrastajúcim napätím znižujú svoje hodnoty najmä v prípade histogramov H_{qmax}, H_{qn}. To znamená, že charakteristiky H_{qmax}, H_{qn} sú nakláňané do pravej strany.

Na základe Obr. 5 je možné konštatovať, že hodnoty šikmosti so vzrastom napätia vo väčšine prípadov zvyšujú svoje hodnoty, taktiež najmä v prípade histogramov H_{qmax}, H_{qn}. To znamená, že charakteristika sa stáva ostrejšou.



Obr. 5 Hodnoty špicatosti vykonaných meraní v závislosti od napätia

V. ZÁVER

Štatistickým spracovaním nameraných údajov je výrazne redukovaná množina vstupných hodnôt do expertného systému, čo umožňuje podstatne zvýšiť rýchlosť klasifikácie vzorov čiastkových výbojov. Okrem štatistickej analýzy bola sledovaná závislosť vybraných popisných operátorov - šikmosť, špicatosť, od aplikovaného napätia. Pre vyššie hodnoty napätia (povrchová výbojová činnosť) sú hodnoty šikmosti H_{qmax} a H_{qn} blízke nule a záporné hodnoty, čo poukazuje na charakteristiku zhodnú s normálovým rozdelením a charakteristiku orientovanú do pravej strany. Pri nižších hodnotách napätia (vnútorná výbojová činnosť a zmiešaná vnútorná a povrchová výbojová činnosť) je charakteristika orientovaná do ľavej strany. Z pohľadu špicatosti pre H_{qmax} a H_{qn}, pri vyšších hodnotách napätia sú aj hodnoty špicatosti vyššie - charakteristika je ostrejšia, pri nižších hodnotách napätia, nižšie hodnoty špicatosti - charakteristika je plochejšia.

POĎAKOVANIE

Tento článok bol vypracovaný s podporou projektu Vývoj unikátneho nízkoenergetického statického zdroja pre elektrosystémy, ITMS 26220220029, ktorý je spolufinancovaný zo štrukturálneho fondu EÚ ERDF v rámci výzvy OPV a V-2008/2.2/01-SORO a prioritnej osi 2 Podpora výskumu a vývoja.

LITERATÚRA

[1] KOLCUNOVÁ, Iraida: Diagnostika elektrických strojov, Technická univerzita Košice 2006, ISBN 80-8073-550-6.
[2] Yazici B.. Statistical Pattern Analysis of Partial Discharge Measurements for Quality Assessment of Insulation Systems in High Voltage Electrical Machinery. SDEMPED 2003, Symposium on Diagnostics for Electric Machines, Power Electronics and Driver, Atlanta, Georgia, GA USA 24-26, August 2003.}.
[3] CIGRE WG 21,03, \"Recognition of Discharges\' Electra, Vol. 11, pp. 61-68,1969.
[4] Sahoo N. C., Salama M. M. A, Bartnikas R.. Trends in Partial Discharge Pattern Classification: A Survey. Sahoo et.al: Trends in Partial Discharge Pattern Classification. IEEE Transactions on Dielectrics and Electrical Insulation, Vol. 12, No. 2, April 2005, pp. 248 - 264. ISSN 1070-9878.
[5] Koktavý P.: Nedestruktívni diagnostika a fyzika dielektrík. Vysoké učení technické v Brne 2006.

Meranie meteorologických parametrov pri využívaní slnečnej energie

Peter KOLPASKÝ, Ján TKÁČ

Katedra elektroenergetiky, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

peter.kolpasky@student.tuke.sk, jan.tkac@tuke.sk

***Abstract* — The presented work deals with the use of renewable energy sources with attention the use of solar energy and its methods measurement. Measuring the intensity of solar radiation is in progress since 1890. During this time were significantly improved methods and instrumentation, and also were amended monitored variables. The most important data on the solar energy is its intensity, which is measured in W/m^2. For this purpose, are currently used devices called Pyranometers Which working mostly on the thermoelectric principle. Very important is the temperature and meteorological parameters during the tests, as well as the practical use of solar installations. This work focuses on meteorological measuring variables related to the use of solar energy through the latest measurement equipment and software LabVIEW. Using renewable energy is consistent with the aim of the EU.**

Keywords **— Solar energy, Pyranometer, temperature, LabVIEW, wind speed, wind direction**

I. Úvod

Jedným zo závažných globálnych a energetických problémov je nedostatok klasických fosílnych palív ako aj vysoká produkcia CO_2 spôsobená aj ich nadmerným využívaním. Riešením tohto vážneho problému je postupný prechod na obnoviteľné zdroje energie. Medzi tieto zdroje patrí v prvom rade slnečná a geotermálna energia.

Tematika obnoviteľných zdrojov a ich využiteľnosti je v poslednom období veľmi frekventovaná vo vedeckých kruhoch ako aj v širokej verejnosti. Globalizáciou a nadmerným využívaním fosílnych palív ľudstvo zapríčiňuje to, že tendencia globálneho otepľovania sa čoraz viac zväčšuje. Aby sa táto tendencia naďalej nezvyšovala, ľudstvo by malo upriamiť svoju pozornosť na využívanie energie, ktoré sú poskytované obnoviteľnými zdrojmi.

Obnoviteľné zdroje, akými sú slnečná energia, vietor, voda, biomasa atď. poskytujú perspektívu, ktorou by ľudstvo v budúcnosti nemalo opovrhnúť.

S pokrokom technologických poznatkov a so zvyšujúcimi sa nákladmi na výrobu elektrickej energie je nutné poznať aké sú alternatívne možnosti na výrobu energie vo forme či už tepla, alebo elektrickej energie. Najčastejšou alternatívou na výrobu týchto foriem energie poslúži energeticky nevyčerpateľný zdroj, ktorým je naše Slnko – slnečná energia.

Slnečná energia má však pri energetickom využívaní svoje špecifiká vyplývajúce z relatívneho pohybu Zeme okolo Slnka. Poskytovaná energie je často premenlivá a na túto premenlivosť vplýva viacero faktorov ako sú ročné obdobia a poveternostné podmienky. Z tohto dôvodu je nutné poznať ako tieto faktory vplývajú na intenzitu slnečného žiarenia. To sa zabezpečuje dlhodobým meraniami parametrov slnečného žiarenia, ako aj meteorologických parametrov v mieste jej využívania. Cieľom tejto práce je realizovať experimentálne merania spomínaných parametrov s využitím rôznych meracích zariadení za účelom merania okamžitých hodnôt ako aj ich dlhodobého sledovania.

II. Meranie Meteorologických Parametrov

Na meranie meteorologických parametrov sa využívajú, buď jednoúčelové špeciálne zariadenia a prístroje, alebo sa merania realizujú zduženými meteorologickými prístrojmi vo forme meteorologických staníc.

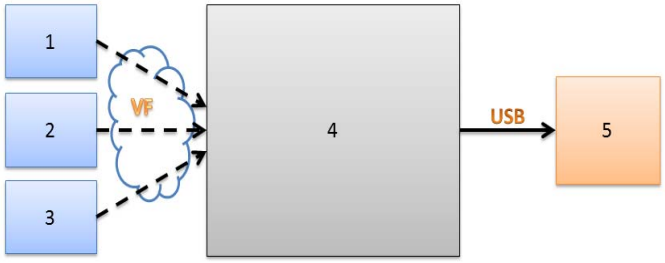
Tieto sú najčastejšie vybavené prístrojmi na meranie týchto veličín:

- Teplota
- Vlhkosť
- Tlak
- Rýchlosť vetra

- Pocitová teplota ochladenia vetrom
- Smer vetra
- Množstvo zrážok...

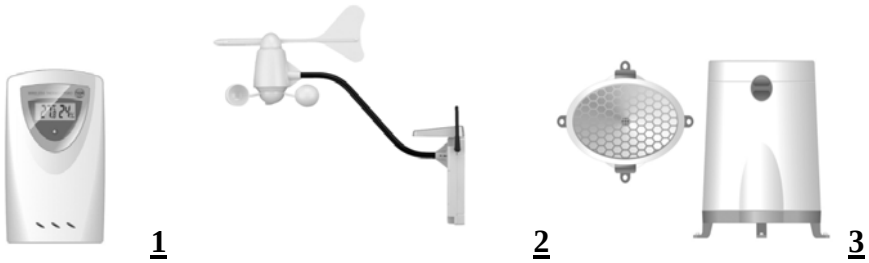
Pri experimentálnych meraniach bola k dispozícií meteorologická stanica podľa blokovej schémy na Obr.1, ktorá pozostáva z:

- Diaľkových senzorov teploty a vlhkosti (termo - hydrometer) -**1**
- Snímača parametrov vetra- rýchlosť a smer - **2**
- Zrážkomeru - **3**
- Centrálnej jednotky - **4**
- Pc -**5**



Obr. 1 Bloková schéma meteorologickej stanice

Konštrukčne je usporiadaná tak, že obsahuje vonkajšie senzory 1, 2, 3, Obr.2 a vnútorný senzor teploty a vlhkosti je umiestnený v centrálnej jednotke Obr.3. Externá jednotka so snímačmi obsahuje aj solárny článok pre nabíjanie batérií, ktoré napájajú vysielač pre prenos informácií. Namerané údaje zo snímačov sú bezdrôtovým spôsobom prenášané do centrálnej jednotky v ktorej sú graficky vyhodnocované a zapisované do internej pamäti, alebo do PC. Centrálna jednotka meteorologickej stanice je na Obr.3 s vyznačenými nasledovnými pozíciami zobrazení jednotlivých veličín 1-tlak vzduchu, 2-teplota a vlhkosť, 3-zrážky, 4-rýchlosť a smer vetra.



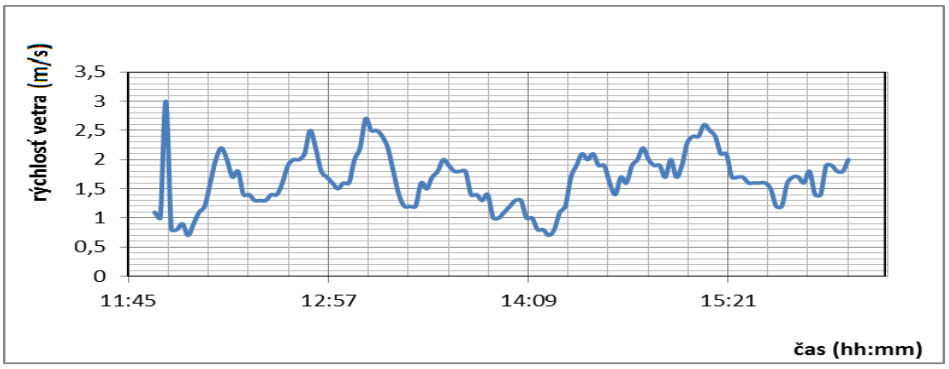
Obr. 1 Snímače meteorologickej *stanice*



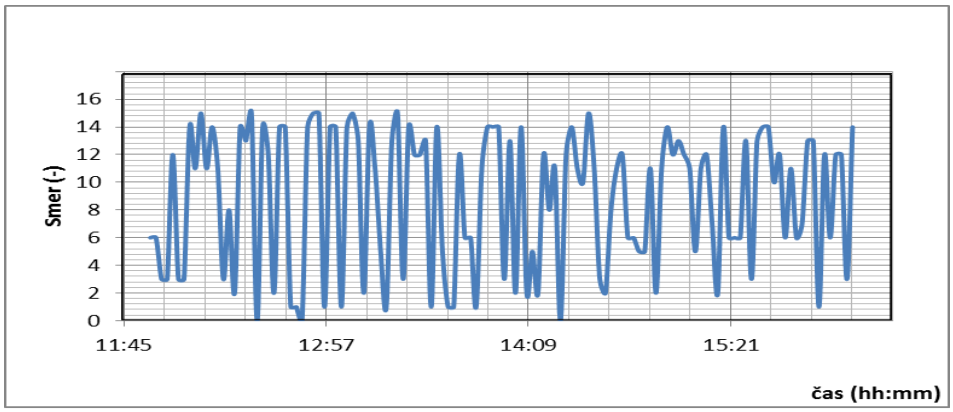
Obr. 3 Centrálna jednotka s displejom

Vyhodnotenie meraných veličín vetra prostredníctvom PC je na Obr4,Obr.5, Obr.6.

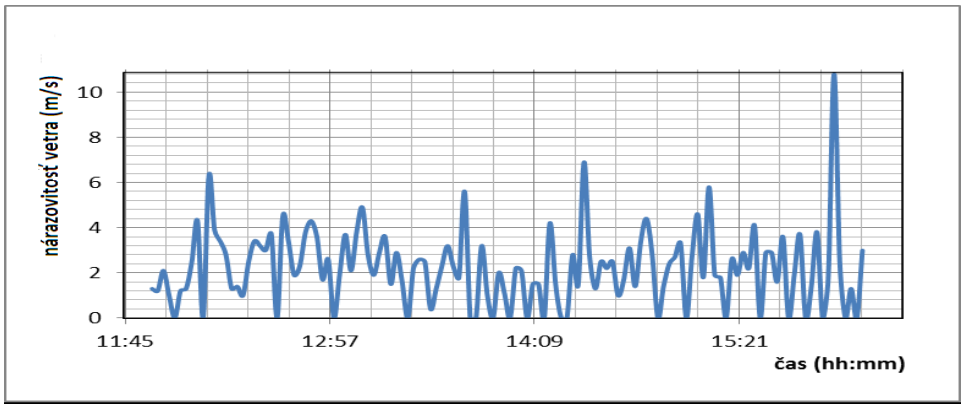
Z uvedených meraní vyplýva, že vietor na stanovišti mal vysokú rýchlostnú a smerovú nestabilitu čomu odpovedala aj vysoká nárazovitosť.



Obr. 4 Priebeh rýchlosti vetra



Obr. 5 Priebeh smeru vetra



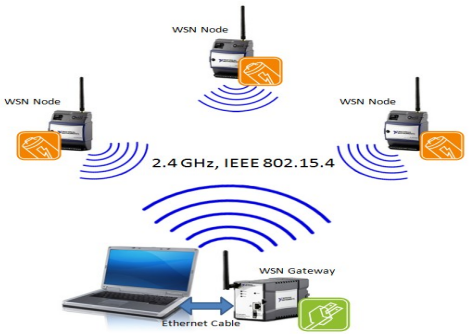
Obr. 6 Priebeh nárazovitosti vetra

Okrem meraní meteorologických parametrov sa merala aj intenzita slnečného žiarenia prichádzajúca zo smeru kolmého k Slnku, globálneho žiarenia na vodorovnú rovinu a difúzneho žiarenia.Meranie bolo realizované pomocou bezdrôtových snímačov WSN podľa zapojenia na Obr. 7.

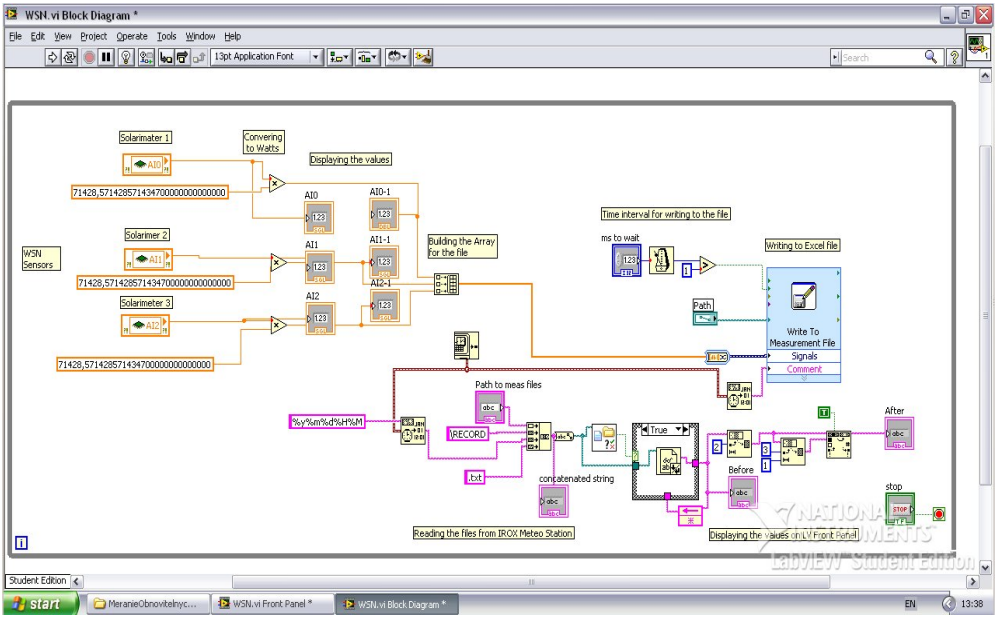
Meracia technika WSN - WIRELESS SENSOR NETWORK (bezdrôtová sieť snímačov) od National Instruments, prináša pokrok v technológii merania z hľadiska toho, že odstránila nutnosť káblového prepojenia s centrálnou jednotkou WSN Gateway.

Ako to vyplýva, z Obr.7 obsahuje sieť snímačov (nody) komunikujúcich navzájom cez bezdrôtový prenos a brány (centrálna jednotka), ktorá namerané data poskytuje prostrednícvom Eternetového výstupu do PC. Namerané údaje boli spravcovávané prostredníctvom programového vybavenia Lab VIEW Grafický program na meranie v prostredí LabVIEW je na

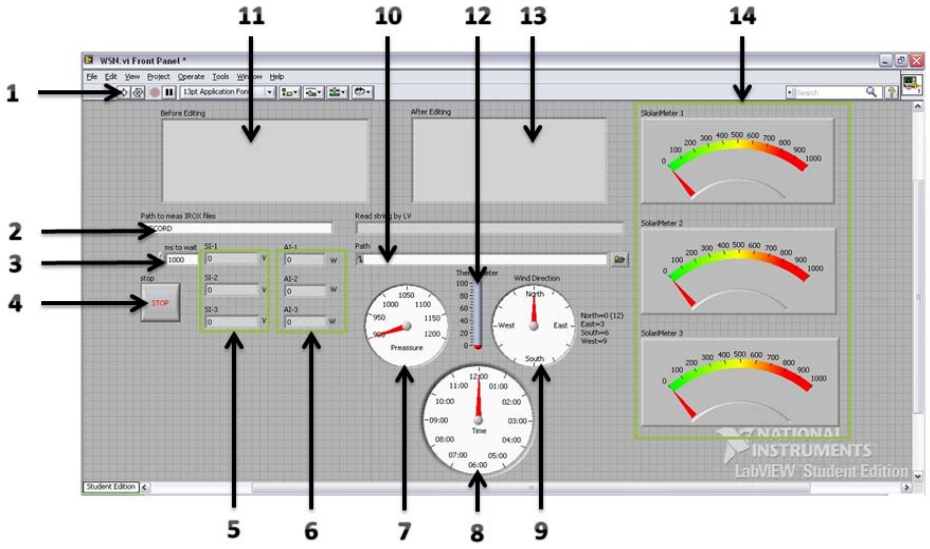
Obr 8. Na Obr.9 je návrh meracieho panelu pre grafické zobrazenie nameraných výsledkov.



Obr. 7 Principiálna schema zapojenia meracej aparatúry WSN



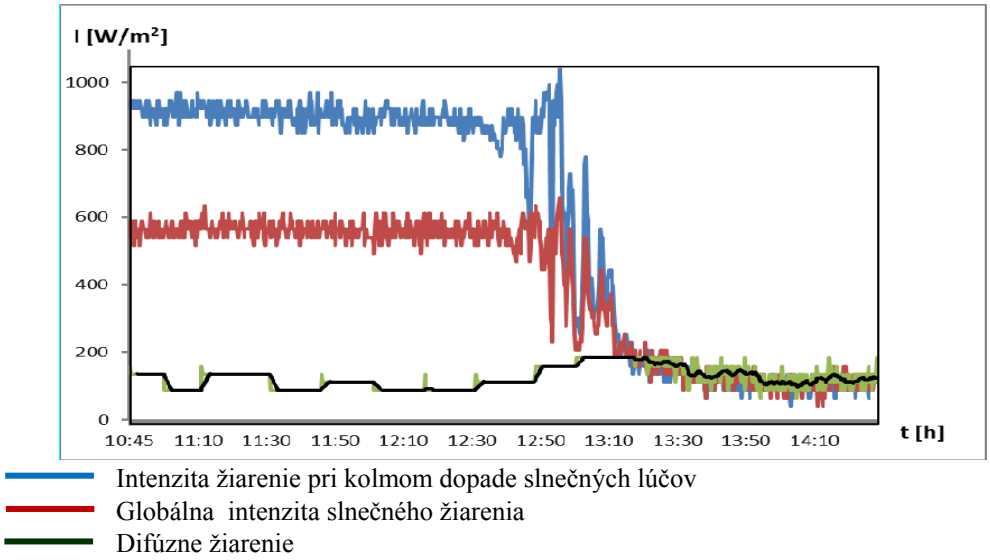
Obr. 8 Program merania v prostredí Lab



Obr.9 Grafické zobrazenie nameraných výsledkov

Na meracom paneli sa zobrazujú okamžité hodnoty a priebehy nameraných hodnôt.
Popis meracieho panelu:
1. – Štart merania
2. – Miesto, odkiaľ sa majú načítavať údaje z meteorologickej stanice

3. – Interval čítania z meteorologickej stanice
4. – Stop merania
5. – Napätia namerané na pyranometroch
6. – Intenzita slnečného žiarenia
7. – Tlak vzduchu získaný z údajov meteorologickej stanice
8. – Čas merania
9. – Smer vetra získaný z údajov meteorologickej stanice
10. –Miesto ukladania zaznamenaných údajov získané z pyranometrov
11. –Údaje načítané z meteorologickej stanice v neupravenej forme
12. –Teplota vzduchu získaná z údajov meteorologickej stanice
13. –Údaje načítané z meteorologickej stanice, neupravenej forme
14. –Intenzita slnečného žiarenia jednotlivých pyranometrov



Obr.10 Výsledky merania intenzity slnečného žiarenia

Na Obr. 10 na ktorom sú graficky vyhodnotené výsledky merania intenzity slnečného žiarenia pomocou troch pyranometrov, z ktorých jeden bol neustále polohovaný kolmo k slnku, druhý zaznamenával globálnu intenzitu slnečného žiarenia na vodorovnú rovinu a tretí pyranometer meral veľkosť zložky difúzneho žiarenia. Počas merania bol zaznamenaný stav, keď dochádzalo k zmene meteorologickej situácie z úplne jasnej oblohy na zamračené. Je zaznamenaný postupný pokles intenzity priameho slnečného žiarenia až na úroveň difúzneho žiarenia. Namerané výsledky potvrdzujú správnosť zvolenej metódy meracej aparatúry ako aj nameraných výsledkov.

III. ZÁVER

Slnečná energia je hlavným a nevyhnutným zdrojom energie pre existenciu života na Zemi. Prechodu slnečného žiarenia cez atmosféru Zeme sa mení na iné formy energie. Túto energiu poskytovanú Slnkom je možné využi pre naše účely na výrobu tepelnej alebo elektrickej energie. Na tieto účely sa vyvíjajú technické prostriedky (fototermálne kolektory, fotovoltické články a atď.).

Na možnosti využitia slnečnej energie majú veľký vplyv poveternostné podmienky. Najčastejšími sledovanými parametrami sú parametre vetra (jeho rýchlosť, smer, pocitové ochladenie a nárazovitosť) , búrková aktivita (množstvo zrážok) a teplota vzduchu. Tieto sa zaznamenávajú prostredníctvom zariadení, ktoré sa nazývajú meteorologické stanice.

Pri zisťovaní množstva energie vyžarovanú Slnkom sa používajú prístroje nazývané pyranometre. Tieto zariadenia, ktoré sa používali pri meraniach premieňali slnečnú energiu na elektrický signál (napätie), ktoré bolo možné zaznamenávať prostredníctvom zariadení, ktoré sú vhodné na meranie signálu veľmi nízkej hodnoty.

Na meranie sa používalo programové prostredie LabVIEW. Tento program získané údaje z meteorologickej stanice z WSN systému prostredníctvom grafického programu spracoval, a následne zobrazil na zobrazovacích jednotkách. Program LabVIEW taktiež načítaval údaje

z meteorologickej stanice a kombináciou týchto veličín je možné porovnať ako vplývajú meteorologické parametre na intenzitu slnečného žiarenia.

Z praktických meraní sa dá usúdiť, že najviac ovplyvňuje intenzitu slnečného žiarenia obsah vlhkosti v atmosfére a výskyt rôznych druhov oblačnosti.

POĎAKOVANIE

Táto práca bola podporovaná Agentúrou na podporu výskumu a vývoja prostredníctvom finančnej podpory projektu č. APVV-0185-10

LITERATÚRA

[1] J.Beranovský, a kol. OBNOVITELNÉ ZDROJE ENERGIE: Základní polovodičové materiály pro solární články. FCC PUBLIC 2001. ISBN 80-901985-8-9
[2] Oneston Solar, photovoltaic technology [online] [cit 2012-3-28]. Dostupné na internete: <http://www.onestonesolar.com/view.php?nid=162>
[3] Enerpro.sro. Solárna mapa Slovenska [online] [cit 2012-4-12]. Dostupné na internete:<http://www.enerpro.sk/?p=22>
[4] What Is a Solarimeter? [online]. Dostupné na internete:<http://www.wisegeek.com/what-is-a-solarimeter.htm>
[5] Katalóg výrobkov National Instruments [online]. Dostupné na internete:<http://www.ni.com>

Komunikácia digitálnych ochrán pomocou protokolu IEC 61850

Andrej MASAROVIČ, Marek HVIZDOŠ

Katedra elektroenergetiky, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

andrej.masarovic@student.tuke.sk, marek.hvizdos@tuke.sk

***Abstrakt* — Tento príspevok popisuje problematiku digitálnych ochrán v elektrických staniciach. Uvádza rôzne komunikačné protokoly používané v praxi. Podrobnejšie rozoberá a charakterizuje medzinárodný štandard IEC 61850. Analyzuje v ňom logický model, typy správ, mapovanie na reálne protokoly a konfiguračný jazyk rozvodne. Demonštruje konfigurovanie generovania správ GOOSE, ktoré si digitálne ochrany posielajú. Tieto správy boli sledované pomocou programu MMS Ethereal, a tým bola potvrdená funkčnosť komunikácie.**

***Kľúčové slová* — digitálna ochrana, IEC 61850, GOOSE, komunikačný protokol**

I. Úvod

Vývoj ochrán spôsobil to, že tieto zariadenia v súčasnosti okrem chránenia dokážu merať, riadiť a monitorovať. Aby si ochrany mohli posielať dáta, je potrebná komunikácia medzi jednotlivými zariadeniami, či už v rámci rozvodne alebo mimo nej.

Vo vývoji komunikácie bolo potrebné prispôsobiť protokoly, pričom vzniklo viacero protokolov pre komunikáciu zariadení v elektrickej stanici. Výsledkom riešenia tejto problematiky je dlhodobo vyvíjaný komunikačný štandard IEC 61850, ktorý sa v súčasnosti presadzuje najviac. Súčasťou tejto normy sú aj tzv. GOOSE správy, ktoré si môžu inteligentné elektronické zariadenia (IED) navzájom posielať.

II. Vývoj ochrán

Postupným vývojom digitalizácie sa aj úroveň ochrán dostala postupne na vyššiu úroveň. Vývojový priebeh ochrán prebiehal v troch etapách. Najprv sa začali používať elektromechanické, nasledovali elektronické a poslednú etapu tvoria digitálne ochrany (terminály).

Digitálne ochrany sa skladajú z elektronických prvkov, ktoré transformujú analógové veličiny do číslicovej formy. Činnosť a logika ochrany je daná obslužným výpočtovým programom. Vzhľadom na možnosti, ktoré softvérový program poskytuje, takéto ochrany poskytujú značný komfort pre obsluhu. Digitálne ochrany majú zabudovanú autokontrolu svojej činnosti, čím je zabezpečený vysoký stupeň spoľahlivosti a predlžuje čas revízie ochrán.

Hlavnou úlohou terminálu ostáva chránenie elektrického zariadenia pri poruche. V bežnej prevádzke zabezpečuje prevádzkové ovládanie jednotlivých prvkov poľa, ich monitorovanie, zber dát o stave poľa a meranie. Automaticky zabezpečuje prenos týchto dát do staničného riadiaceho informačného systému (RIS) a následne na dispečing.

III. Komunikačné protokoly

Elektrické zariadenia v elektrickej stanici dokážu vďaka komunikačným protokolom medzi sebou komunikovať. Komunikačný protokol je súbor syntaktických a sémantických pravidiel určujúcich výmenu informácií medzi dvoma entitami spojenými napr. prostredníctvom počítačovej siete. Protokol riadi ako sa správy obsahujúce dáta a informácie zhromažďujú v zdroji pre ich prenos cez sieť k cieľu. Medzi tieto komunikačné protokoly patria napríklad DNP 3.0, Modbus, Profibus, Alstom Courier, Elcom-90, IEC 60870-5, IEC 61850, atď.

A. DNP 3.0 Protokol

DNP bol vytvorený v roku 1990. V roku 1993 bol sprístupnený celosvetovo a akceptovaný užívateľskými skupinami. V roku 1995 bol znova vylepšený na verziu DNP 3.0. DNP 3.0 je otvorený inteligentný, robustný a efektívny moderný SCADA protokol.

Vlastnosti protokolu:

- Môže odosielať a prijať viacnásobné dátové typy v jednej správe
- Môže odpovedať bez požiadavky
- Podporuje časovú synchronizáciu a štandardný časový formát
- Zahŕňa iba zmenu dát v odpovedajúcej správe

B. Modbus RTU Protokol

Modbus RTU protokol podporuje obe Master a Slave služby. Slave protokol poskytuje spôsob, ako komunikovať s Modbus Master a naopak. Modbus podporuje až štyri Modbus rozhrania naraz.

Nasledujúce Modbus kódy sú podporované (v závislosti na použití Master alebo Slave):

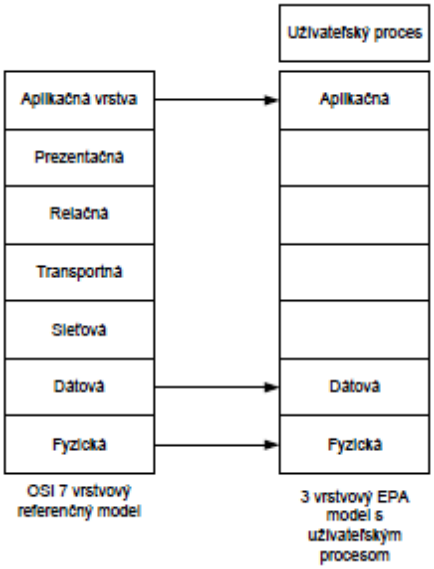
- Čítanie vstupných stavov
- Čítanie výstupných registrov
- Čítanie vlastných registrov
- Nastavenie registra
- Stanovenie viacnásobného registra

C. ABB – INDACTIC 2033 Protokol

Protokol INDACTIC je vlastníctvom ABB a je využívaný na výmenu dát medzi RTU a riadiacim strediskom (Control Center). RTU a Control Center sú časti SCADA PN 40 systému. INDACTIC 2033 vytvorila Veesta Automation správajúca sa ako ABB-RTU umožňujúca spracovať dáta a príkazy z riadiacej stanice. Tento protokol je dodávaný pod prísnym licenčným dohľadom dodávateľa ABB.

D. IEC 60870-5

Protokolový súbor IEC 60870-5 bol vyvinutý komisiou IEEE v roku 1995, definujúcou komunikáciu a radenie telemetrických systémov SCADA (Supervisory Control and Data Acquisition) v energetických sieťach. Architektúra tohto protokolového súboru vychádza z OSI referenčného modelu, ktorý je redukovaný na tri vrstvy. Na Obr. 1 je štruktúra OSI modelu s trojvrstvovým EPA (Enhanced Performance Architecture) modelom. Ide o OSI model bez sieťovej, transportnej, relačnej a prezentačnej vrstvy. Je teda možné tvrdiť, že celý systém je závislý na riadiacom zariadení.



Obr. 1 Porovnanie OSI modelu s 3 vrstvovým EPA modelom

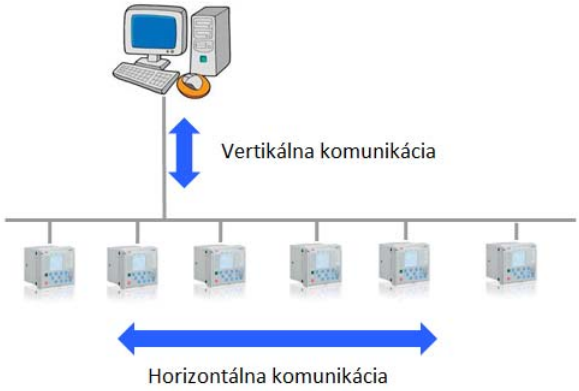
IV. SPÔSOBY KOMUNIKÁCIE

A. Vertikálna komunikácia

Vertikálna komunikácia je časovo nekritická komunikácia medzi zariadeniami.. Počas komunikácie sa využíva metóda prístupu Klient/Server. Ochrany predstavujú server a dávajú svoje dáta k dispozícii klientom. Väčšinou je možné pripojiť ku každej ochrane až päť klientov súčasne. Nevýhodu časovej nekritickosti odstraňuje tým, že umožňuje aj klientskym staniciam aby riadili prenos dát prostredníctvom horizontálnej komunikácie, čím prináša výpočtový výkon bližšie k procesu.

B. Horizontálna komunikácia

Je používaná pre rýchlu horizontálnu komunikáciu medzi zariadeniami. Jediná takáto správa môže byť prijatá viacerými zariadeniami a je obrazom priradených MMS (Manufacturing Message Specification) dát. Môže byť používaná pre zdieľanie analógových hodnôt, indikáciu pozícií, vzájomné blokády a k najdôležitejším patrí možnosť prenášať kritické vypínacie signály.



Obr. 2 Spôsob komunikácie

V. MEDZINÁRODNÁ NORMA IEC 61850

Vo svete sa pre komunikáciu v rozvodniach používa veľké množstvo protokolov, ale len nový pohľad na stanice združuje komplexný súbor IEC 61850. Predstavuje jednotnú, štandardizovanú, na dodávateľovi nezávislú metódu pre tvorbu komunikačnej siete a integrácií jednotlivých zariadení rozvodni. Základným cieľom pri tvorbe tohto súboru noriem bolo umožniť vytváranie systému, v ktorých budú komunikovať zariadenia od rôznych výrobcov. Tieto zariadenia, spojené komunikačnou sieťou sa označujú skratkou IED, ktoré zaisťujú ochranu rozvodní, kontrolu nad ich prevádzkou a automatizáciou, meraniami a reguláciou v rozvodni.

Norma IEC 61850 obsahuje desať hlavných častí. Za primárnu časť sa považujú časti šesť až desať. V nasledujúcej tabuľke je popísaný obsah štandardu IEC 61850.

Tabuľka 1
Obsah normy IEC 61850

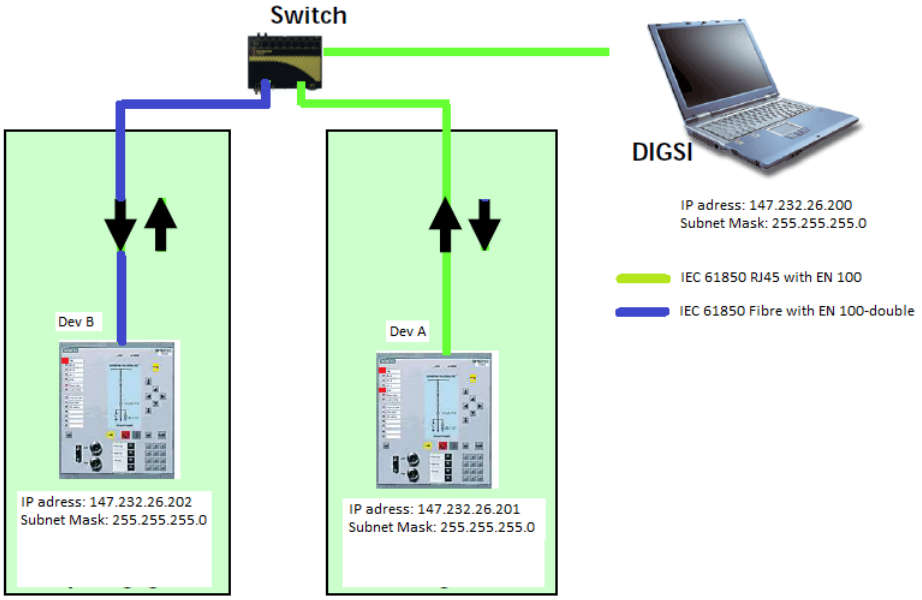
Číslo časti	Názov
IEC 61850-1	Úvod a prehľad
IEC 61850-2	Slovník
IEC 61850-3	Všeobecné požiadavky
IEC 61850-4	Systémové a projektové riadenie
IEC 61850-5	Komunikačné požiadavky na funkcie a modely prístrojov
IEC 61850-6	Jazyk na opis konfigurácie na komunikáciu v staniciach s inteligentnými elektronickými zariadeniami (IED)
IEC 61850-7-1	Základná komunikačná štruktúra. Zásady a modely
IEC 61850-7-2	Základné informácie a komunikačná štruktúra. Abstraktné rozhranie komunikačnej služby (ACSI)
IEC 61850-7-3	Základná komunikačná štruktúra. Všeobecné triedy údajov
IEC 61850-7-4	Základná komunikačná štruktúra. Kompatibilné triedy logických uzlov a triedy údajov uzlov a dátové triedy
IEC 61850-7-410	Vodné elektrárne. Komunikácia pri monitorovaní a riadení
IEC 61850-7-420	Logické uzly zdrojov distribuovania energie
IEC 61850-8-1	Mapovanie špecifickej komunikačnej služby (SCSM). Mapovanie podľa MMS (ISO 9506-1 a ISO 9506-2) a podľa ISO/IEC 8802-32) a ISO/IEC 8802-3
IEC 61850-9-1	Špecifické mapovanie komunikačných služieb (SCSM) - prenos vzorkových hodnôt po sériovom jednosmernom viacbodovom spoji Bod-Bod
IEC 61850-9-2	Mapovanie špecifickej komunikačnej služby (SCSM). Vzorové hodnoty vyššie ako ISO/IEC 8802-3viacbodového spojenia Bod-Bod
IEC 61850-10	Skúšanie zhody

VI. KOMUNIKÁCIA DIGITÁLNYCH OCHRÁN SIPRTOREC

Praktická časť je realizovaná prostredníctvom dvoch ochrán typu Siemens SIPROTEC, ktoré vzájomne komunikujú prostredníctvom GOOSE telegramov.

Postup pri komunikácií je realizovaný v týchto krokoch:

1. Stlačenie funkčného tlačidla F1 na zariadení DevA – vykoná obsluha
2. Rozsvietenie LED diódy (LED1) na prednom paneli ochrany DevA
3. Poslanie GOOSE správy z ochrany DevA na zariadenie DevB
4. Prijatie GOOSE správy v ochrane DevB
5. Rozsvietenie LED diódy na zariadení DevB (LED1)
6. Poslanie GOOSE správ zo zariadenia DevB na DevA, ktoré potvrdzujú prijatie pôvodnej správy
7. Prijatie GOOSE správy ochranou DevA od zariadenia DevB
8. Rozsvietenie LED diódy (LED5) na zariadení DevA
9. Stlačenie funkčného tlačidla F2 na DevA – vykoná obsluha
10. Zhasnutie LED diód na zariadeniach DevA a DevB

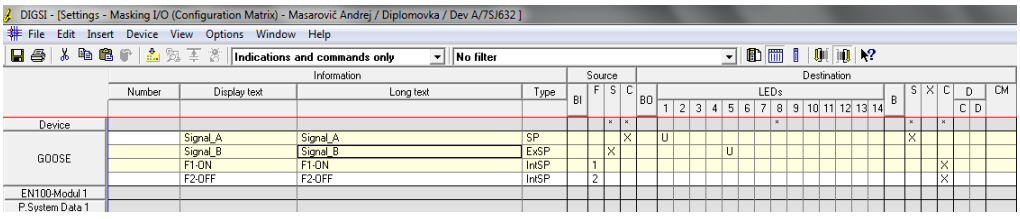


Obr. 3 Posielania GOOSE správ zariadení DevA a DevB

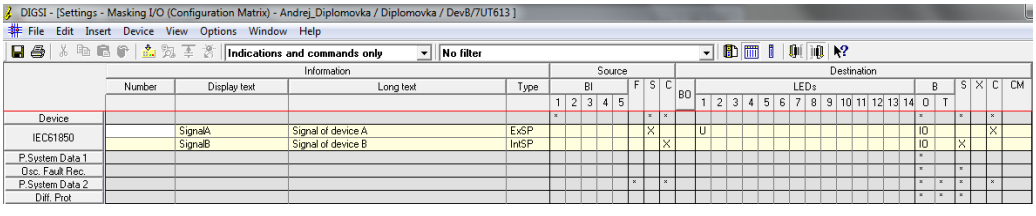
A. Nastavenie ochrán pomocou programu DIGSI

Pre správne fungovanie komunikácie GOOSE správ je potrebné vykonať nastavenie v niekoľkých krokoch. Prvotné je nastavenie IP adries. Po tomto nastavení je potrebné vytvoriť v Masking I/O signály, ktoré ochrany budú vysielať a prijímať (**Obr. 4** a **Obr. 5**).

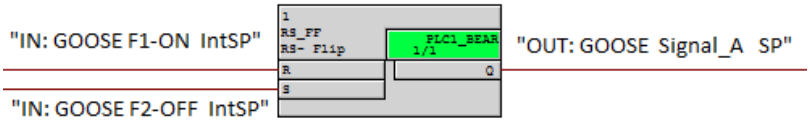
Aby tieto signály boli vôbec funkčné je potrebné ich vytvoriť aj v CFC logike ochrán (**Obr. 6** a **Obr. 7**).



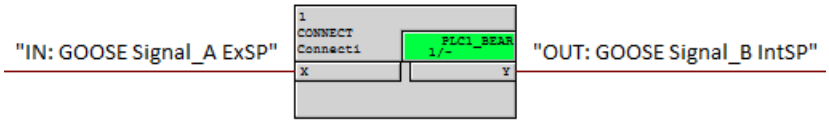
Obr. 4 Nastavenie signálov na ochrane Dev A v Masking I/O



Obr. 5 Nastavenie signálov na ochrane Dev B v Masking I/O



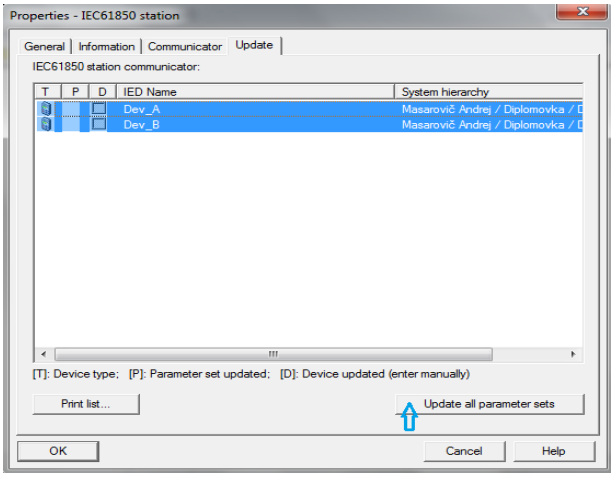
Obr. 6 Vytvorenie CFC logiky na ochrane Dev A



Obr. 7 Vytvorenie CFC logiky na ochrane Dev B

B. Vytvorenie IEC 61850 Station

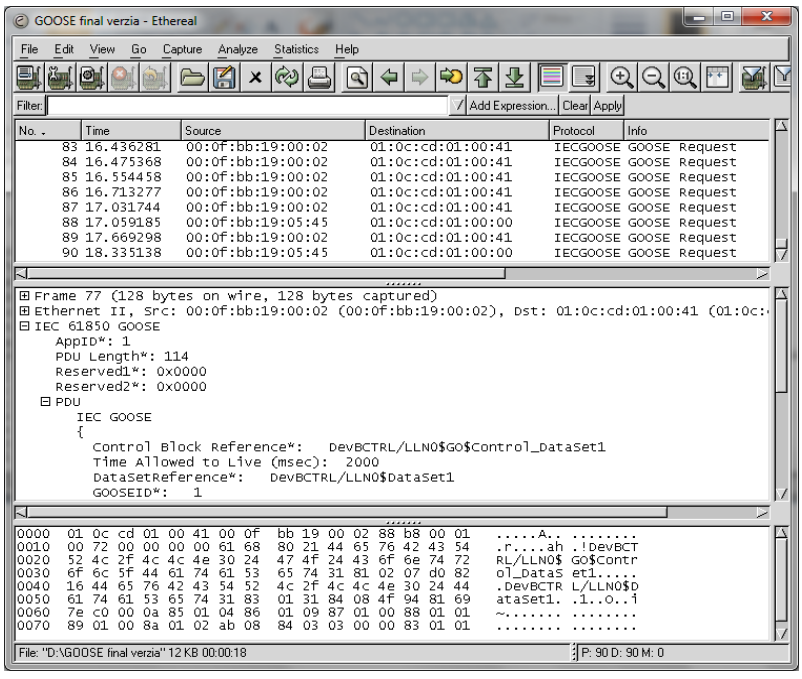
IEC 61850 je položkou v programe DIGSI, ktorá sa pridáva do projektu. Takáto stanica spája niekoľko zariadení komunikujúcich medzi sebou prostredníctvom Ethernetu v súlade s IEC 61850. V IEC 61850 sa nastavuje to, ktorá ochrana bude ako zdroj GOOSE správ a ktorá bude príjemca. Po následnom nastavení sa ochrany aktualizujú (**Obr. 8**).



Obr. 8 Aktualizovanie ochrán

C. Sledovanie GOOSE správ pomocou programu MMS Ethereal

Na sledovanie GOOSE správ existuje viacero programov. V tomto prípade je použitý MMS Ethereal (**Obr. 9**).



Obr. 9 Sledovanie správ GOOSE

Stručný popis zaznamenaného hexadecimálneho kódu GOOSE správy v programe MMS Ethereal je uvedený v Tab. 2.

Tabuľka 2
Popis hexadecimálneho kódu

No.	GOOSE správa	Popis
1	00 0c cd 01 00 00	Cieľová adresa pre GOOSE
2	00 of bb 19 05 45	Zdrojová adresa pre ochranu
3	88 b8 01	Typ: GOOSE IEC 61850
4	00 01	Identifikačné číslo aplikácie v tomto prípade je 1
5	00 72	Dĺžka GOOSE správy je 72 bitov
6	00 00	Vyhradené 1
7	00 00	Vyhradené 2
8	61 68	Implicitná sekvencia, dĺžka PDU je 104 bitov
9	80 21	GOOSE riadiaci blok, 9 bitov
	4c 4c 4e 30 24 47 4f 24 43	LLN0$GO (Logický uzol GOOSE riadiaceho bloku)
10	81 02	Čas existencie GOOSE, 2 bite, max. 20 ms

VII. ZÁVER

Tento článok poukazuje na využitie protokolu IEC 61850 pri komunikácii zariadení v elektrických staniciach. Jeho využitie v praxi sa začína čoraz viac presadzovať, práve pre jeho jednotnosť pre všetky zariadenia od rôznych svetových výrobcov. Výhody GOOSE správ spočívajú v tom, že nahradzujú konvenčné metódy komunikácie medzi tzv. inteligentnými elektronickými zariadeniami. Pri komunikácii medzi týmito zariadeniami sa v súčasnosti čoraz viac využívajú telegramy GOOSE aj kvôli ich spoľahlivosti, ktorá je vysoká vďaka ich opakovanému vysielaniu.

LITERATÚRA

[1] F. Janíček – V. Chladný – A. Beláň – Ž. Eleschová: *Digitálne ochrany v elektrizačnej sústave*. Bratislava, 2004, ISBN 80-227-2135-2.

[2] Veesta World: Realtime Software development system Communication Protocols[online]. Veesta World 2002-2005. [cit. 2012-05-03]. Dostupné na internete: <http://www.veesta-world.com/pages/services_protocol_page.htm>

[3] K. Kulski – D. Tholomier: Impact of IEC 61850 on substation automation and protection. In: *Proceedings of the Fourth International Scientific Symposium ELEKTROENERGETIKA 2007*, Technical University of Košice, 2007, p. 505-510, ISBN 978-80-8073-844-0.

Využitie nízkopotenciálneho tepla pomocou tepelných čerpadiel

Peter MASKA, Marek HVIZDOŠ, Ján TKÁČ

Katedra elektroenergetiky, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

peter.maskal@student.tuke.sk, marek.hvizdos@tuke.sk, jan.tkac@tuke.sk

***Abstrakt* — Predkladaný príspevok sa zaoberá problematikou využívania obnoviteľných zdrojov energie so zameraním na využívanie tepelných čerpadiel. Tepelné čerpadlá boli zaradené medzi obnoviteľné zdroje vzhľadom na vysoký kvantitatívny potenciál nízkopotenciálneho a odpadového tepla. Tieto zariadenia dokážu využívať nízkopotenciálne teplo a transformovať ho na teplo s vyšším potenciálom, ktoré je znova použiteľné. Prispievajú tým k úspore energie a znižovaniu spotreby. V príspevku je popísaný princíp činnosti jednotlivých druhov tepelných čerpadiel. Pozornosť je venovaná aj reverzibilným tepelným čerpadlám, ktoré sú celoročne využiteľné. Skonštruované laboratórne tepelné čerpadlo vzduch-voda, na ktorom boli vykonané laboratórne merania, dosahuje parametre sériovo vyrábaných tepelných čerpadiel.**

***Kľúčové slová* — tepelné čerpadlo, alternatívne zdroje energie, nízkopotenciálne teplo**

I. Úvod

Z dôvodu znižovania zásob vyčerpateľných zdrojov energie a poškodzovania životného prostredia je v súčasnosti veľký záujem o energeticky efektívne a ekologické zariadenia. Medzi tieto zariadenia patria obnoviteľné zdroje energie využívajúce solárnu, veternú, vodnú energiu a energiu biomasy.

Predkladaný príspevok je zameraný na tepelné čerpadlá, ktoré využívajú solárnu energiu premenenú na teplo obsiahnuté vo vode, zemi a vzduchu. Taktiež využívajú teplo zo zemského jadra, ktoré odoberajú z hornín a geotermálnych prameňov. Výhodou tepelných čerpadiel je schopnosť odoberať nízkopotenciálne teplo, ktoré je prakticky nevyužiteľné a premieňať ho na teplo s vyšším potenciálom.

Cieľom predkladaného príspevku je poukázať na výhody použitia tepelných čerpadiel, ktoré predstavujú modernú alternatívu vykurovania a chladenia.

Ďalšia časť príspevku je orientovaná na skonštruované laboratórne tepelné čerpadlo vzduch-voda, na ktorom boli vykonané laboratórne merania. Účelom laboratórneho merania je zistenie zmeny teploty na jednotlivých častiach tepelného čerpadla, teploty média vo vyhrievanej nádrži a spotrebu elektrickej energie pri zohrievaní média. Z merania je určené výkonové číslo, podľa ktorého je možné zistiť efektívnosť tepelného čerpadla.

II. Princíp činnosti tepelných čerpadiel

Úlohou tepelných čerpadiel je odoberať nízkopotenciálne teplo z okolitého prostredia a transformovať ho na teplo s vyšším potenciálom. Podľa spôsobu získavania a odovzdávania tepla a podľa spôsobu zabezpečovania pohonnej energie existuje viacero druhov tepelných čerpadiel. Jednotlivé druhy kladú rôzne požiadavky na prevádzku, inštaláciu a konštrukciu.

A. Rozdelenie tepelných čerpadiel podľa pohonu

Tepelné čerpadlá sa z hľadiska pohonu rozdeľujú na kompresorové a absorpčné. U kompresorových tepelných čerpadiel zabezpečuje pohon kompresor a pri absorpčných tepelných čerpadlách je to vysokoteplotná tepelná energia.

Kompresorové tepelné čerpadlá

Odparovaním média sa privádza nízkopotenciálne teplo z vonkajšieho prostredia. Vyparovanie sa uskutočňuje na primárnej strane tepelného čerpadla vo výmenníku tepla

(výparník). V druhej polovici výparníka sa cez expanzný ventil pod veľkým tlakom vstrekuje médium. V dôsledku nižšieho tlaku vo výparníku sa médium ľahko odparuje. Teplota média vo výparníku je menšia ako teplota prostredia z ktorého sa odoberá teplo. Ohriate médium vo forme plynu je odsávané kompresorom, v ktorom sa stláča, a tým sa zvyšuje jeho tlak a teplota. Pary média sú ďalej odovzdávané do sekundárneho výmenníka (kondenzátora), v ktorom sa teplo odvedie do vykurovacieho okruhu alebo do zariadenia na ohrev teplej úžitkovej vody. Médium je znova vedené do expanzného ventilu, v ktorom expanduje, pričom sa prudko zníži jeho teplota a tlak. Médium opäť zmení svoje skupenstvo na kvapalné. Celý cyklus sa následne opakuje. Pohon kompresora zabezpečuje najčastejšie elektromotor. Ďalšou alternatívou je spaľovací motor alebo aj Stirlingov motor, ktorý je poháňaný teplom [1].

Absorpčné tepelné čerpadlá

Pracujú na podobnom princípe ako kompresorové tepelné čerpadlá. Kompresia pracovnej kvapaliny u absorpčných tepelných čerpadlách je uskutočnená v obvode, ktorý sa skladá z absorbéru, čerpadla roztoku, generátora a expanzného ventila. Para s nízkym tlakom vychádzajúca z výparníka je pohlcovaná absorbentom, pričom vzniká teplo. Roztok pri prechode cez čerpadlo roztoku nadobudne vyšší tlak. Roztok ďalej vstupuje do generátora, v ktorom sa pracovná kvapalina odparuje pri vysokej teplote, za pomoci externého zdroja. V kondenzátore dochádza k skvapalneniu pracovnej kvapaliny, pričom absorbent sa vracia cez expanzný ventil do absorbéru. Vo výparníku sa z tepelného zdroja odoberá teplo. Užitočné získané teplo je odovzdávane cez kondenzátor a taktiež aj absorbér. V generátore sa teplo s vysokou teplotou používa na pohon celého procesu [2].

B. Rozdelenie tepelných čerpadiel podľa spôsobu získavania a odovzdávania tepla

Pri označovaní druhov tepelných čerpadiel sa používajú výrazy vzduch, zem a voda. Prvý výraz označuje odkiaľ tepelné čerpadlo odoberá tepelnú energiu (zem, voda, vzduch) a druhý výraz označuje ako tepelné čerpadlo dodáva energiu do objektu (voda, vzduch) [3].

Tepelné čerpadlá zem-voda

Patria medzi významné systémy a vzhľadom na prevádzkovanie voči vonkajším klimatickým podmienkam sa považujú za najstabilnejšie. Väčšinou sa prevádzkujú v bivalentnom chode, čiže pri nízkej teplote sa k tepelnému čerpadlu pripája zdroj tepla.

Teplo sa zo zeme odoberá buď pomocou zemných kolektorov alebo zemných sond. Zemné kolektory sa ukladajú horizontálne do zeme do hĺbky cca 1,5 až 2 m. Na ich uloženie je potrebný dostatočne veľký priestor. Zemná sonda je omnoho drahšia ako zemný kolektor. Zaberá menej miesta, je uložená vertikálne a umiestňuje sa až do hĺbky 150 m [4].

Tepelné čerpadla zem-voda majú vďaka dobrej stabilite systému veľmi dlhú životnosť. Vďaka nezávislosti voči vonkajšej teplote sa môžu používať počas celého roka. Ich nevýhodou sú vysoké investičné náklady a náročné zemné práce pri inštalácii.

Tepelné čerpadlá vzduch-voda

Tepelné čerpadlá vzduch-voda sa skladajú z vnútornej a vonkajšej jednotky, no existujú aj kompaktné vyhotovenia. Výkon tepelného čerpadla závisí od vonkajšej teploty vzduchu.

Tento systém je výhodný z hľadiska jednoduchej inštalácie a veľkej univerzálnosti. Dá sa nainštalovať na akýkoľvek typ stavby. Nevýhodou je hlučnosť spôsobená prechodom vzduchu a potreba bivalentnej prevádzky pri nízkych teplotách vonkajšieho vzduchu.

Tepelné čerpadlá voda-voda

Teplo získavajú z povrchovej alebo podzemnej vody. V oblasti s vhodnými geologickými podmienkami s výdatným prameňom je najlepším zdrojom tepelnej energie studňa. Teplota podzemnej vody je pomerne stabilná a pohybuje sa okolo 10 °C. Tepelné čerpadlo potrebuje pre svoju funkciu dve studne, pričom jedna je zdrojová a druhá nasávacia. Odporúča sa umiestniť studne vo vzájomnej vzdialenosti minimálne 15 m. Pri bežnom rodinnom dome by mala byť výdatnosť prameňa aspoň 0,5 l/s.

Tento systém dosahuje najlepšie hodnoty výkonového čísla. Nevýhodou je málo vhodných oblastí pre jeho inštaláciu a kolísanie teploty vody povrchových tokov. Prevádzkové náklady sú kvôli potrebe čerpania vyššie.

Tepelné čerpadlá vzduch-vzduch

Pracujú na podobnom princípe ako tepelné čerpadlá vzduch-voda. Tepelná energia je na

sekundárnej strane predávaná prostredníctvom vzduchu.

Ich výhodou je jednoduchá inštalácia a nižšie ceny. Nevýhodou je hlučnosť a nižšia efektívnosť v porovnaní s inými druhmi tepelných čerpadiel [3].

C. Reverzibilné tepelné čerpadlá

Reverzibilné tepelné čerpadlá sa okrem vykurovania a ohrevu teplej úžitkovej vody používajú aj na chladenie. Pre tento účel sa najčastejšie používajú tepelné čerpadlá druhu vzduch-vzduch alebo vzduch-voda pracujúce v monovalentnej alebo v bivalentnej prevádzke.

V prípade reverzibilného tepelného čerpadla vzduch-voda pracuje tepelné čerpadlo v chladnejších obdobiach vo vykurovacom režime. Tepelné čerpadlo odoberá teplo z vonkajšieho vzduchu a využíva ho na ohrev vody v radiátoroch.

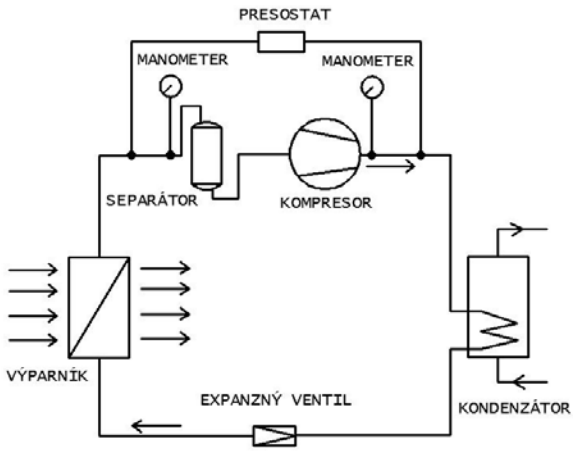
Počas letnej sezóny pracuje reverzibilné tepelné čerpadlo v chladiacom režime. Počas tejto prevádzky bude kondenzátor plniť úlohu výparníka. Tým sa bude z miestnosti odoberať teplo prostredníctvom radiátorov, ktoré sa môže ďalej využiť na ohrev teplej úžitkovej vody [5].

III. MERANIE NA LABORATÓRNOM TEPELNOM ČERPADLE

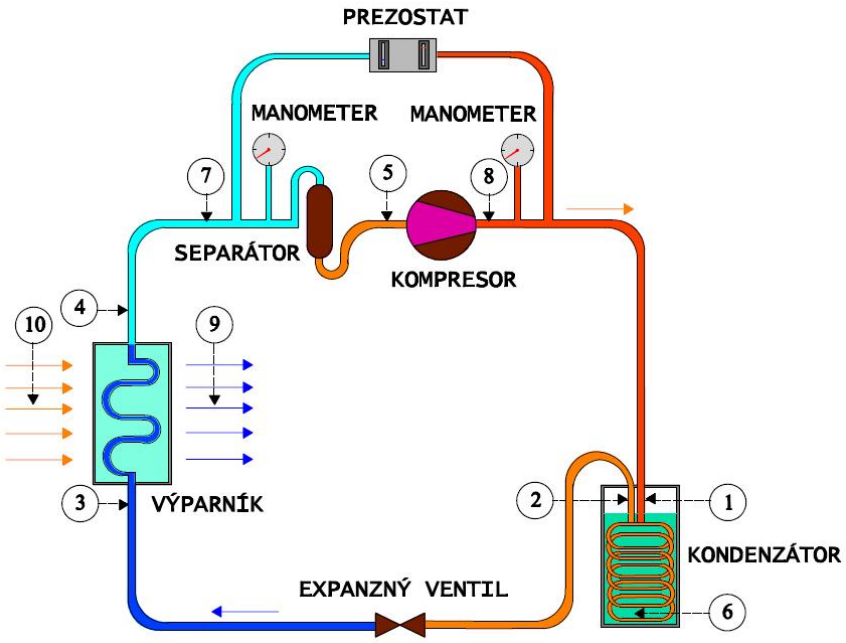
Pre overenie efektívnosti tepelného čerpadla bolo skonštruované laboratórne tepelné čerpadlo vzduch-voda, ktoré odoberalo tepelnú energiu z okolia a ohrievalo médium v nádrži (Obr. 1). Principiálna schéma tepelného čerpadla je znázornená na Obr. 2.



Obr. 1 Laboratórne tepelné čerpadlo vzduch-voda



Obr. 2 Principiálna schéma laboratórneho tepelného čerpadla

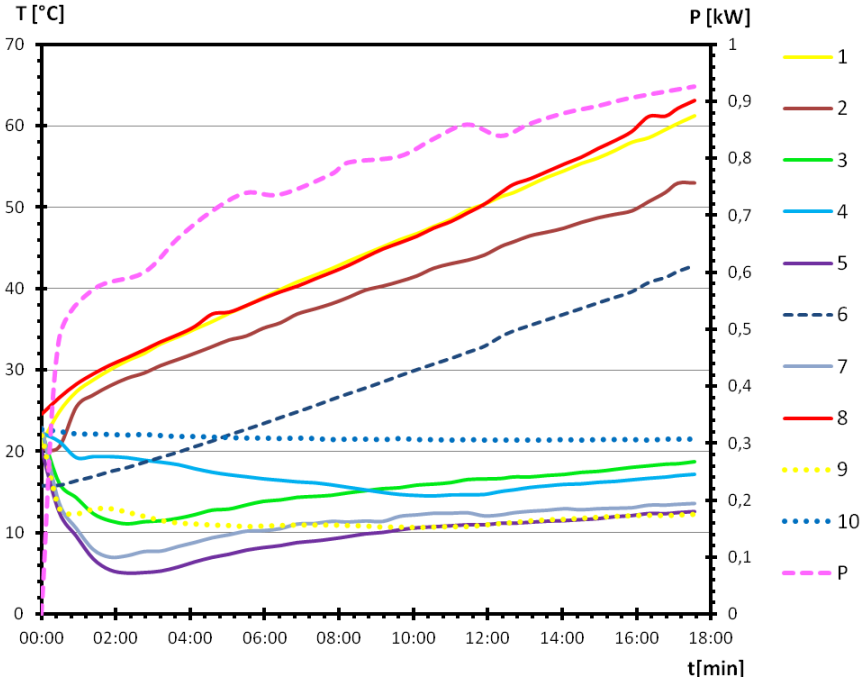


Obr. 3 Umiestnenie teplotných snímačov na laboratórnom tepelnom čerpadle

Prostredníctvom teplotných snímačov 1 až 10 sa merala teplota ohrievaného média, teplota nasávaného a vyfukovaného vzduchu a teplota jednotlivých častí tepelného čerpadla. Umiestnenie teplotných snímačov je na Obr. 3. Signál z teplotných snímačov bol prevedený cez merací systém do počítača. Počas merania bola zaznamenávaná spotreba elektrickej energie potrebnej na pohon tepelného čerpadla. Teplota okolitého vzduchu bola určená jednak snímačom teploty nasávaného vzduchu a taktiež prostredníctvom externého teplomera.

Popis funkcie jednotlivých snímačov:

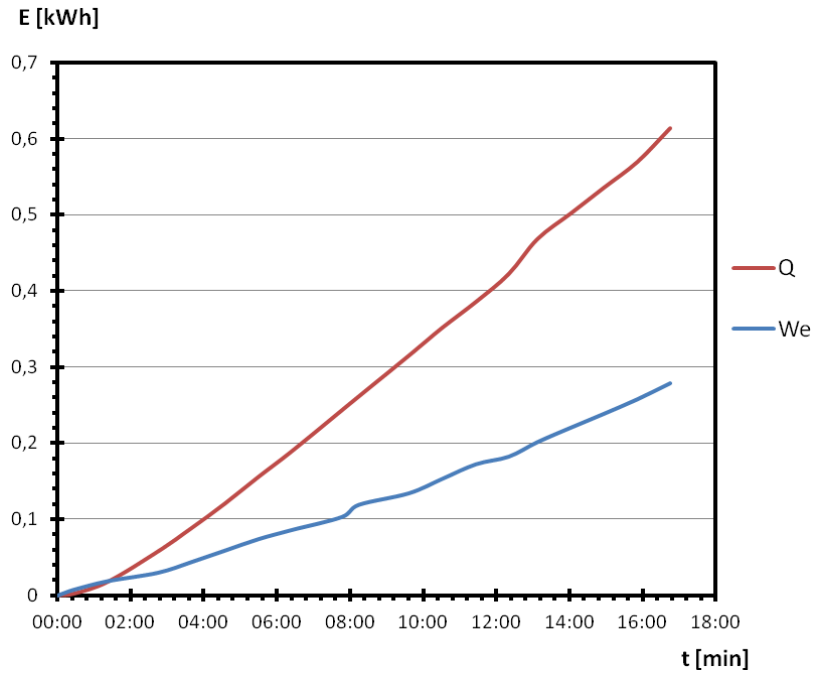
1 – teplota na vstupe do kondenzátora
2 – teplota na výstupe z kondenzátora
3 – teplota na vstupe do výparníka
4 – teplota na výstupe z výparníka
5 – teplota na vstupe do kompresora
6 – teplota ohrievaného média
7 – teplota na vstupe do separátora
8 – teplota na výstupe z kompresora
9 – teplota vyfukovaného vzduchu
10 – teplota nasávaného vzduchu



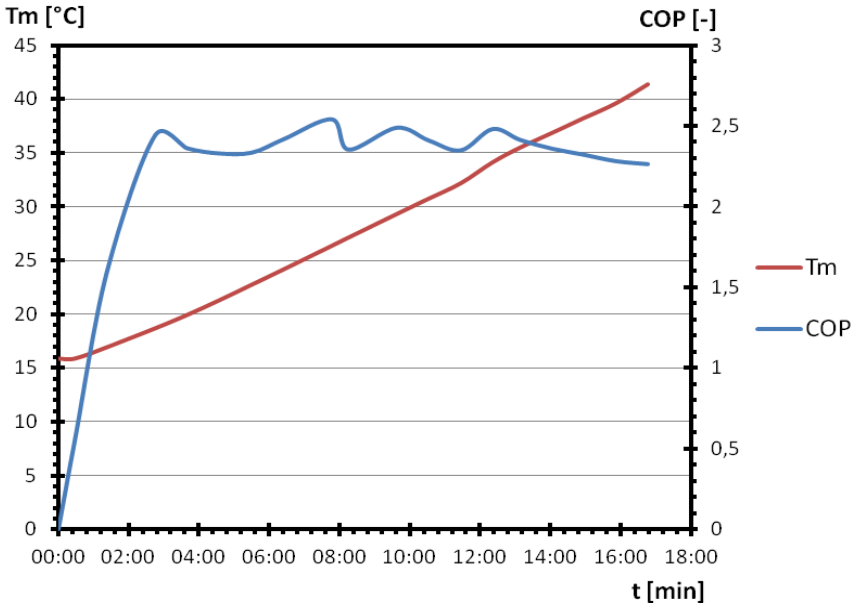
Obr. 4 Zmeny teplôt na tepelnom čerpadle a príkon tepelného čerpadla počas merania

Meranie prebiehalo pri teplote okolia cca 22 °C. Na začiatku merania malo médium v nádrži s množstvom 21 l teplotu 15,81 °C. Za 18 minút sa médium zohrialo na 41,38 °C. Počas ohrievania média v nádrži dochádzalo k zmenám teplôt na tepelnom čerpadle a postupne sa zvyšovala hodnota príkonu tepelného čerpadla (Obr. 4).

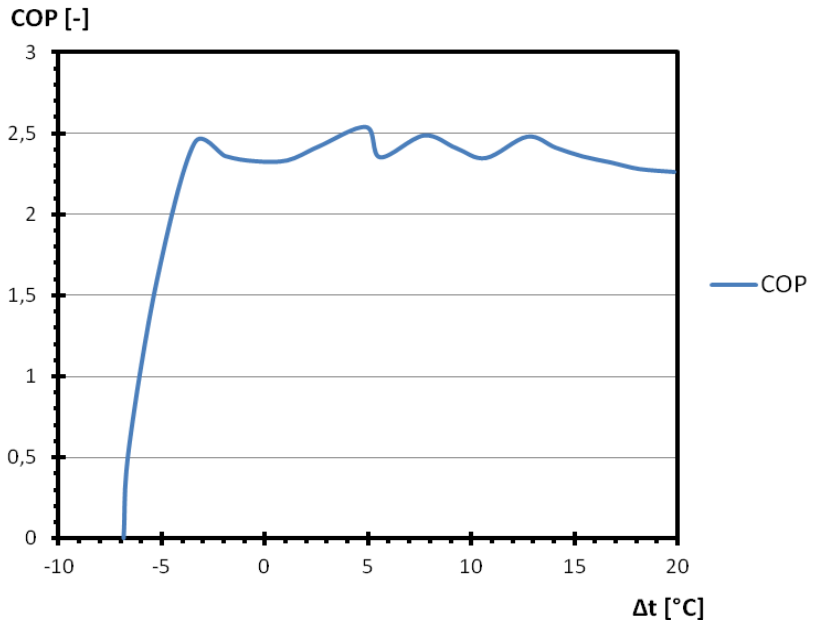
Výkonové číslo tepelného čerpadla (*COP*) sa pohybovalo v rozmedzí hodnôt 2,2 až 2,5. Z tohto vyplýva, že tepelné čerpadlo vyrobilo viac ako dvojnásobok spotrebovanej elektrickej energie. Nárast vyrobenej tepelnej energie oproti spotrebovanej elektrickej energie je znázornený na Obr. 5. Na začiatku merania začali hodnoty výkonového čísla prudko narastať v dôsledku nízkej teploty ohrievaného média. Zohrievaním média v nádrži začalo výkonové číslo mierne klesať (Obr. 6). Výkonové číslo sa menilo aj v závislosti od teplotného spádu medzi vonkajším prostredím a ohrievaným médiom (Obr. 7).



Obr. 5 Nárast vyrobenej tepelnej energie oproti spotrebovanej elektrickej energii počas merania



Obr. 6 Zmena *COP* pri zohrievaní média v nádrži



Obr. 7 Závislosť *COP* od teplotného spádu medzi okolitým prostredím a ohrievaným médiom

IV. ZÁVER

Súčasné tepelné čerpadlá sa oproti starším typom zdokonalili tak, že okrem vykurovania objektov a ohrev teplej úžitkovej vody sa dokážu použiť aj na chladenie priestorov. Vďaka tejto funkcii sa tepelné čerpadlá môžu používať počas celého roka, a tým ich uplatnenie nadobudlo ešte väčší význam.

Pri voľbe vykurovacieho zariadenia objektu je potrebné zvážiť všetky možnosti a vybrať zariadenie, ktoré je z hľadiska nákladov na prevádzku a efektívnosti najvýhodnejšie. Napriek vysokej dobe návratnosti tepelného čerpadla sa tieto zariadenia oplatí využívať. Doba návratnosti sa môže znížiť prostredníctvom štátnej podpory a dotáciami alebo zavedením špeciálnych sadzieb na tepelné čerpadlá.

V Laboratóriu NZE bolo skonštruované laboratórne tepelné čerpadlo vzduch-voda, odoberajúce tepelnú energiu z vonkajšieho prostredia, ktorá sa využila na ohrev teplonosného média. Na tepelnom čerpadle boli vykonané merania pri rôznych teplotách vonkajšieho vzduchu. Meranie potvrdilo, že účinnosť tepelného čerpadla sa mení aj podľa veľkosti teplotného spádu. Výsledky merania poukazujú na efektívnosť a úspornosť tepelných čerpadiel. Laboratórne tepelné čerpadlo umožňuje tepelný zisk minimálne vo výške dvojnásobku vloženej elektrickej energie. Namerané parametre dosahujú úrovne sériovo vyrábaných tepelných čerpadiel.

POĎAKOVANIE

Táto práca bola podporovaná Agentúrou na podporu výskumu a vývoja prostredníctvom finančnej podpory projektu č. APVV-0185-10.

Vďaka za podporu projektu VaV operačného programu, Centrum excelentnosti výkonových elektronických systémov a materiálov pre ich komponenty II. No. OPVaV-2009/2.1/02-SORO, ITMS 26220120046. Projekt je financovaný zo zdrojov ES, ERDF – Európsky fond regionálneho rozvoja.

LITERATÚRA

[1] A. Žeravík: *Stavíme tepelné čerpadlo*. Ing. Antonín Žeravík, 2003, 312 p., ISBN 80-239-0275-X.
[2] EU.SK: Tepelné čerpadlo [online]. [citované 2011-09-18]. Dostupné na internete: <http://www.tepelnecerpadla.eu.sk/>.
[3] R. Karlík: *Tepelné čerpadlo pro váš dům*. Grada Publishing, Praha, 2009, 112 p., ISBN 978-80-247-2720-2.
[4] eFilip: *Tepelné čerpadlo* [online]. [citované 2011-06-09]. Dostupné na internete: <http://www.e-filip.sk/default.aspx?contentID=1273>.
[5] Glen Dimplex Deutschland: *Reverzibilné tepelné čerpadlo* [online]. [citované 2012-02-06]. Dostupné na internete: <http://www.energline.sk/files/tc-reverzibilne.swf>.

Modelovanie elektromagnetických polí v okolí stožiarov vvn a zvn

Dušan MEDVEĎ, Jozef ŠURIN

Katedra elektroenergetiky, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská Republika

dusan.medved@tuke.sk, jozef.surin@student.tuke.sk

***Abstrakt* — Tento príspevok sa zaoberá riešením rozloženia elektromagnetických polí v okolí stožiarov veľmi vysokého a zvlášť vysokého napätia.**

***Kľúčové slová* — elektromagnetické pole, metóda konečných prvkov, ANSYS, vonkajšie silové vedenie**

I. Úvod

Na mechanické upevnenie vodičov vonkajšieho silového vedenia sa používajú stožiare, t.j. konštrukčné časti na „nesenie“ vodičov, upevnené pomocou izolátorových reťazcov, nakoľko vodiče majú iný potenciál, ako je potenciál Zeme. Tvar a rozmery stožiarov závisia od napätia, druhu a počtu vodičov a uzemňovacích lán, veľkosti rozpätí, terénu, nosnosti pôdy, použitého materiálu, funkcie stožiaru a síl, ktorým má stožiar odolávať. Výška stožiara je stanovená na základe miesta umiestnenia. Závisí od členitosti terénu, usporiadania vodičov na hlave stožiara, priehybu vodičov, dĺžky izolátorových reťazcov, vzdialenosti uzemňovacieho lana od vodičov a od predpísanej výšky spodného vodiča od zeme alebo križovaných objektov.

Pre dvojité 110 kV vedenie sa používa väčšinou jednodriekový úzky stožiar s deleným (pätkovým, stienkovým) základom. Pre jednoduché 220 kV a 400 kV sa používal portálový stožiar. V súčasnosti sa portálové a viacdriekové stožiare v SR nesmú používať. Portálové stožiare nahradili stožiare tvaru „Y“. Pre dvojité 220 kV a 400 kV vedenia sa používajú tzv. stožiare typu „Donau“.

Vodičmi vonkajšieho silového vedenia tečú značné prúdy a tak v ich blízkosti, ako aj v blízkosti stožiarov, môže byť elektromagnetické pole (EMP) životu nebezpečné, prípadne môže mať nepriaznivé účinky. Na eliminovanie vplyvu elektromagnetických polí je potrebné poznať správne a presné rozloženie elektromagnetického poľa. V ďalšej časti tohto príspevku budú bližšie popísané výsledky riešenia rozloženia EMP v blízkosti 110 a 400 kV stožiarov.

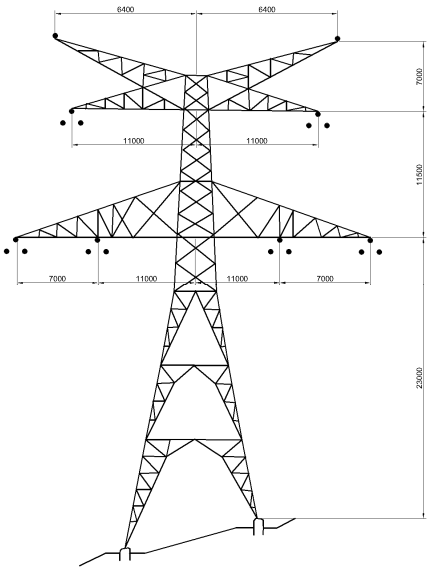
II. Modelovanie Elektromagnetického Poľa v Okolí Eelektroenergetických Zariadení

Pre numerický výpočet rozloženia elektromagnetického poľa sa využívajú Maxwellove rovnice doplnené o materiálové rovnice. Pre ich riešenie je vhodné použiť niektorú numerickú metódu (metódu konečných prvkov (MKP), metódu konečných diferencií (MKD) alebo metódu hraničných prvkov (MHP)). Profesionálne výpočtové programy využívajúce MKP pre riešenie elektromagnetických polí sú pre svoju univerzálnosť vhodné aj pre takýto typ úlohy. Model stožiarov bol zjednodušený do 2D priestoru, čím sa aj skrátil čas výpočtu oproti 3D modelu. Geometrické a materiálové parametre stožiara, upevnenie a typ lán sú uvedené v tabuľkách 1 a 2.

III. Modelovanie Elektromagnetického Poľa v Okolí 400 kV Vedenia

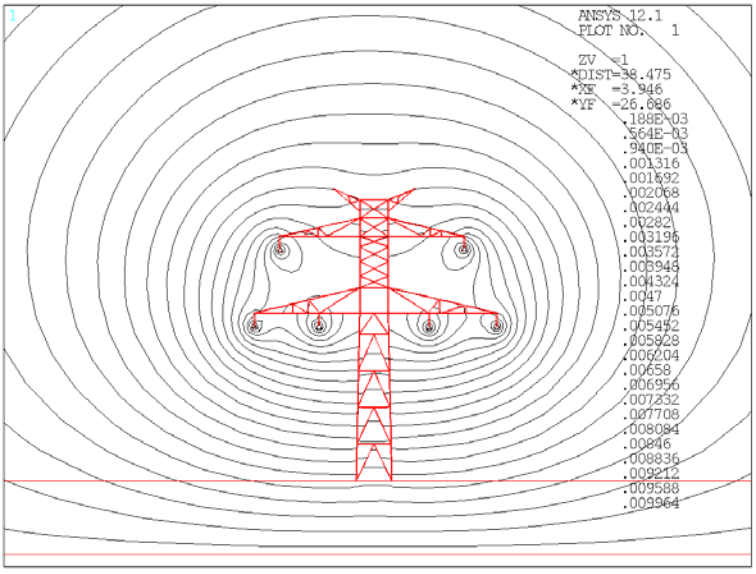
Tabuľka 1
Technické parametre 2×400 kV vedenia typu DONAU

Typ stožiara	Donau
Počet 3f systémov	2
Počet zemných lán	2
Napäťová úroveň	400 kV
Fázový prúd	≤ 3×835 A
Frekvencia siete	50 Hz
Typ fázového vodiča	AlFe 450/52
Typ zemného lana	AlFe 180/59
Počet vodičov vo zväzku	3

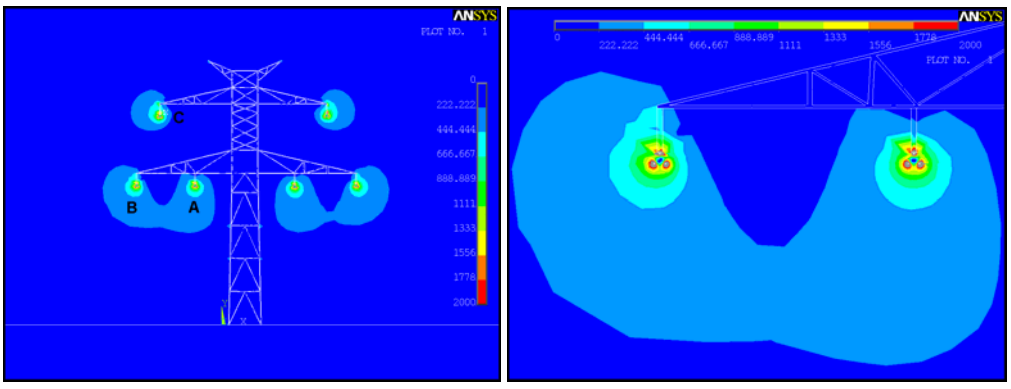


Stožiar typu „Donau“

A. *Výsledky*

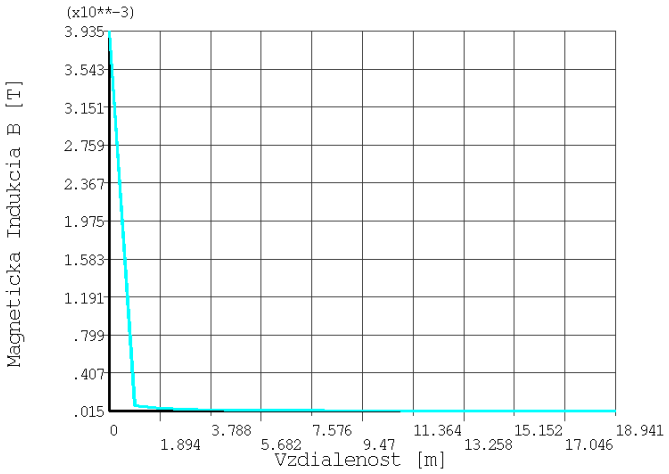


Obr. 1 Rozloženie siločiar magnetického poľa v okolí 400 kV vedenia



a) b)

Obr. 2 a) Rozloženie magnetickej intenzity v okolí 400 kV vedenia; b) Detail rozloženia magnetickej intenzity v okolí zväzkového vodiča 400 kV vedenia (A a B)



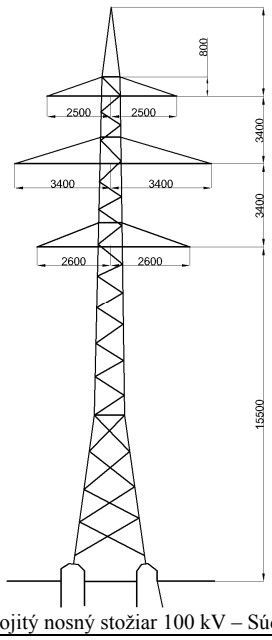
Obr. 3 Grafická závislosť magnetickej indukcie od vzdialenosti (medzi bodom A a zemou (Obr. 2a))

Na (Obr. 3) je znázornená grafická závislosť magnetickej indukcie ***B*** [T] v závislosti od vzdialenosti *d* [m] medzi fázovým vodičom „A“ (Obr. 2a) a zemou. Ako je vidieť z grafu, veľkosť magnetickej indukcie smerom od fázového vodiča „A“ k zemi prudko klesá a vo vzdialenosti približne 20 m od vodiča je hodnota magnetickej indukcie 15 µT . Táto hodnota predstavuje reálnu hodnotu magnetickej indukcie, ktorú je možné namerať priamo pod vedením vysokého napätia. Zákon č. 355/2007 Z.z. upravuje akčné hodnoty expozície hustoty magnetickej indukcie pre frekvenciu 50 Hz, ktorá pre obyvateľstvo je max 100 nT. V tomto prípade teda dochádza k prekročeniu akčnej hodnoty expozície magnetickej indukcie a je potrebné vykonať opatrenia, aby sa zabránilo nežiaducim účinkom na obyvateľstvo. Je potrebné zdôrazniť, že táto zvýšená hodnota ja priamo pod 400 kV vedením. Vo vzdialenosti 90 m od fázového vodiča „B“ dosahuje magnetická indukcia už nepatrné hodnoty. Z tejto simulácie tiež vyplýva, že zdržiavanie sa priamo pod vonkajším silovým vedením môže mať nežiaduci vplyv na ľudský organizmus a môže vážne ohroziť zdravie človeka.

IV. Modelovanie Elektromagnetického Poľa v Okolí 110 kV Vedenia

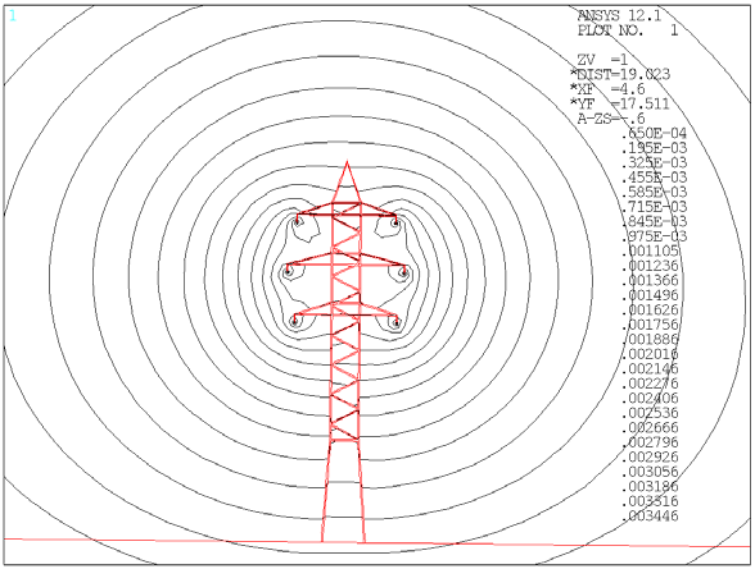
Tabuľka 2
Technické parametre 2×110 kV vedenia typu „Súdok“

Typ stožiara	Jednodriekový úzky
Počet 3f systémov	2
Počet zemných lán	1
Napäťová úroveň	110 kV
Fázový prúd	535 A
Frekvencia siete	50 Hz
Typ fázového vodiča	AlFe 240/39
Typ zemného lana	AlFe 180/59

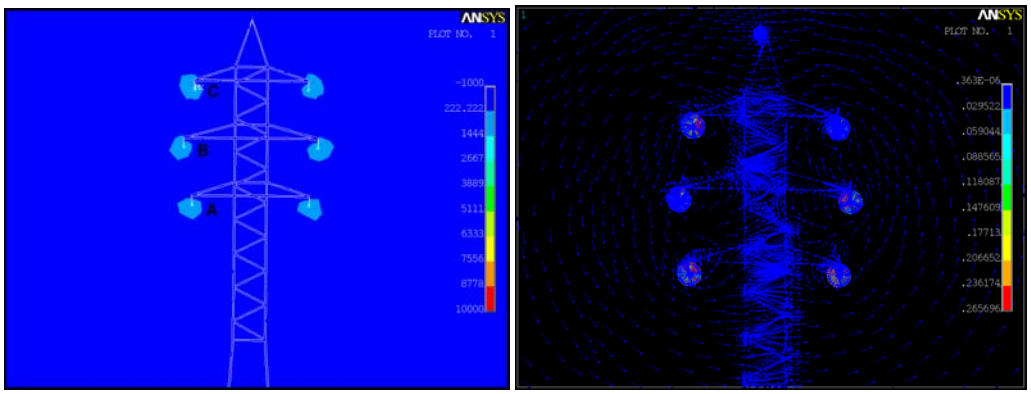


Dvojitý nosný stožiar 100 kV – Súdok

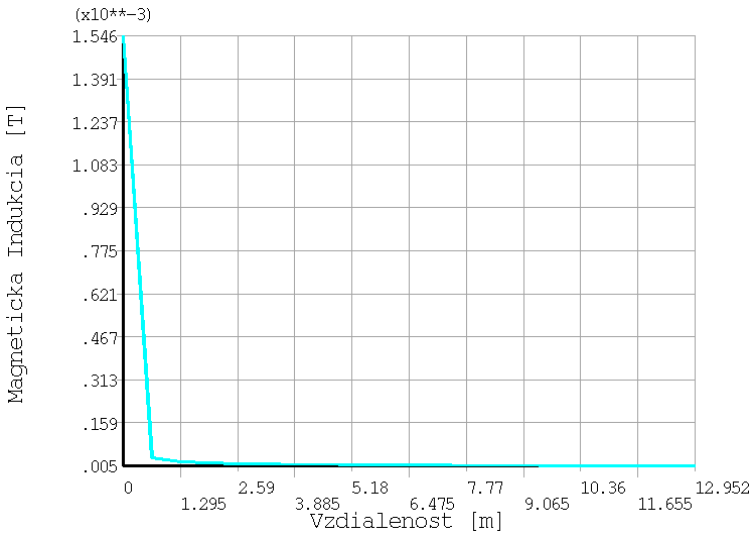
B. Výsledky



Obr. 4 Rozloženie siločiar magnetického poľa v okolí 110 kV vedenia



Obr. 5 a) Rozloženie magnetickej intenzity v okolí 110 kV vedenia; b) Vektorové znázornenie rozloženia magnetickej indukcie v okolí 110 kV vedenia



Obr. 6 Grafická závislosť magnetickej indukcie od vzdialenosti (medzi bodom A a zemou (Obr. 5a))

Na (Obr. 6) je znázornená grafická závislosť magnetickej indukcie ***B*** [T] v závislosti od vzdialenosti *d* [m] medzi fázovým vodičom „A“ (Obr. 5a) a zemou (kolmo na zem). Ako je vidieť z grafu, veľkosť magnetickej indukcie smerom od fázového vodiča „A“ k zemi prudko

klesá a vo vzdialenosti približne 13 m od vodiča fázy „A“ je hodnota magnetickej indukcie približne 5 μT. Aj v tomto prípade ide o prekročenie dovolenej hodnoty magnetickej indukcie, ako v predchádzajúcom prípade (400 kV vedenia). Vo vzdialenosti približne 100 m od fázového vodiča „A“ dosahuje magnetická indukcia už nepatrné hodnoty, ktoré neohrozujú zdravie človeka.

Elektromagnetické pole v miestach pod vonkajším vedením je možné eliminovať, resp. znížiť niekoľkými spôsobmi, napr.:

- zväčšením vzdialenosti fázových vodičov od zeme,
- znížením prevádzkového napätia,
- zaistením kompaktnosti vedenia – riadením vzdialenosti vodičov,
- zmenou sledu fáz u viacsystémových vonkajších vedení,
- pridaním tieniacich lán.

POĎAKOVANIE

Tento príspevok vznikol vďaka podpore v rámci operačného programu Výskum a vývoj pre projekt Ochrana obyvateľstva SR pred účinkami elektromagnetických polí (ITMS kód: 26220220145), spolufinancovaný zo zdrojov Európskeho fondu regionálneho rozvoja.



Podporujeme výskumné aktivity na Slovensku/Projekt je spolufinancovaný zo zdrojov EÚ.

LITERATÚRA

[1] ŠURIN Jozef: *Modelovanie elektromagnetických polí v okolí energetických zariadení*. (Modeling of electromagnetic fields around the power equipment). Diplomová práca. Košice: Technická univerzita v Košiciach, Fakulta elektrotechniky a informatiky, 2012.
[2] VARGA, Ladislav – ILENIN, Stanislav – BALOGH, Jozef: *Elektrické rozvody*, Košice, 2006. 144 s.
[3] DĚDKOVÁ, Jarmila: *Modelování elektromagnetických polí*, VUT, Brno 2005.
[4] NOVÁK, Pavol – MEDVEĎ, Dušan: *Premeny elektrickej energie*. Košice: Technická univerzita. 2012. 194 s. ISBN 978-80-553-0883-8.
[5] NOVÁK, P.: *Elektrické teplo*. Košice. Equilibria. 2010. ISBN978-80-89284-48-1.
[6] ROT, D. *Efektivní řešení fyzikálně sdružených problémů z oblasti elektrotepelných technologií*. Plzeň : Západočeská univerzita, 2009, 165 s.
[7] ROT, D. – KOŽENÝ, J. *Modelování startu tavby oxidů kovů elektromagnetickou indukcí ve studeném kelímku. ElectroScope*, 2010, roč. 2010, č. 4, s. 1-6. ISSN: 1802-4564.
[8] 355/2007 Z. z. – *Zákon o ochrane, podpore a rozvoji verejného zdravia a o zmene a doplnení niektorých zákonov* (čiastka 154/2007).
[9] *Stožiare vonkajších elektrických vedení*: [online] Dostupné na: < http://www.kves.uniza.sk/kvesnew/dokumenty/Mechanika%20veden%C3%AD/Mechanika_10_Stoziare.ppt >.
[10] VOJTEK, M. – LAGO, J. – SAVČÁK, M.: *Hodnotenie úrovní elektrických a magnetických polí "Vedenia 2x400 kV TR Moldava - USS"*. [online] Dostupné na: < http://eia.enviroportal.sk/dokument?id=34314 >.
[11] FRIMER Róbert: *Analýza vplyvu nízkofrekvenčného elektromagnetického žiarenia na vnútornu klímu budov a možnosti ochrany a znižovania expozície*. NOVUS SCIENTIA 2007, 20.11.2007 ÚVZ Herľany, Slovenská republika, ISBN: 978-80-8073-922-5.

Modelovanie prechodných javov využitím nástroja Matlab/SimPowerSystem

Dušan MEDVEĎ, Lukáš NEMERGUT

Katedra elektroenergetiky, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

dusan.medved@tuke.sk, lukas.nemergut@student.tuke.sk

***Abstrakt* —Tento príspevok sa zaoberá modelovaním prechodných javov v elektrizačnej sústave využitím nástroja Matlab/SimPowerSystem.**

***Kľúčové slová* — elektrizačná sústava, prechodný jav, SimPowerSystem**

I. Úvod

Pre správne dimenzovanie elektrických vedení a správne nastavenie elektrických ochrán v elektrizačnej sústave (ES), je nutné poznať skratové pomery prúdov v príslušnom mieste sústave. V skúmanej ES je potrebné určiť maximálne a ustálené hodnoty skratových prúdov. Výpočet spomínaných veličín sa uskutočňuje na základe známych parametrov prvkov siete a tiež zapojenia jednotlivých zariadení – zdrojov, vedení, záťaží. Do výpočtu skratových pomerov sa zahŕňajú všetky napájacie zdroje a príspevky od elektrických točivých strojov.

Výpočet skratových prúdov je založený na riešení diferenciálnych rovníc, prípadne zjednodušených algebrických rovníc – je potrebné poznať impedancie príslušných vetiev obvodu až do miesta skratu a hodnoty napájacích zdrojov. Z týchto údajov sa vypočíta hodnota skratového prúdu v mieste skratu a určia sa príspevky od jednotlivých napájacích zdrojov.

Keďže v trojfázových sústavách vznikajú nesymetrické poruchové stavy, je vhodné systém nelineárnych rovníc zjednodušiť. Nesymetrická trojfázová sústava sa zvyčajne nahradí symetrickými zložkami fázorov súslednej, spätnej a netočivej sústavy.

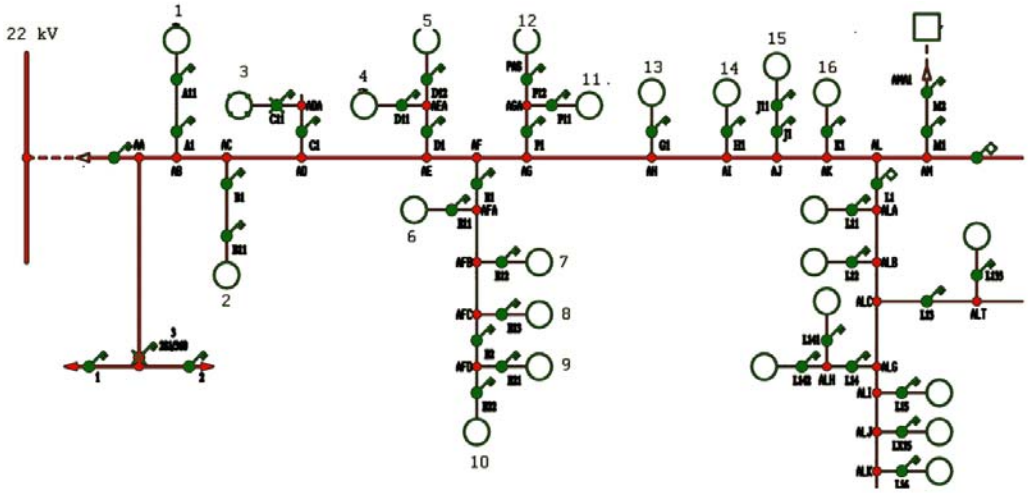
V súčasnej dobe existuje množstvo komerčných aj voľne šíriteľných nástrojov, ktoré spomínanú analýzu uskutočňujú. Jedným z často využívaným je aj Matlab a jeho nástroj Simulink/SimPowerSystem, v ktorom budú uskutočnené aj nasledujúce simulácie prechodných dejov na vybranej časti lúčovej siete (oblasť na východnom Slovensku).

II. Analýza Prechodných Dejov v 22 kV Sieti

V tejto časti budú uvedené výsledky simulácií niektorých vybraných prechodných dejov, ktoré môžu nastať v 22 kV sieti.

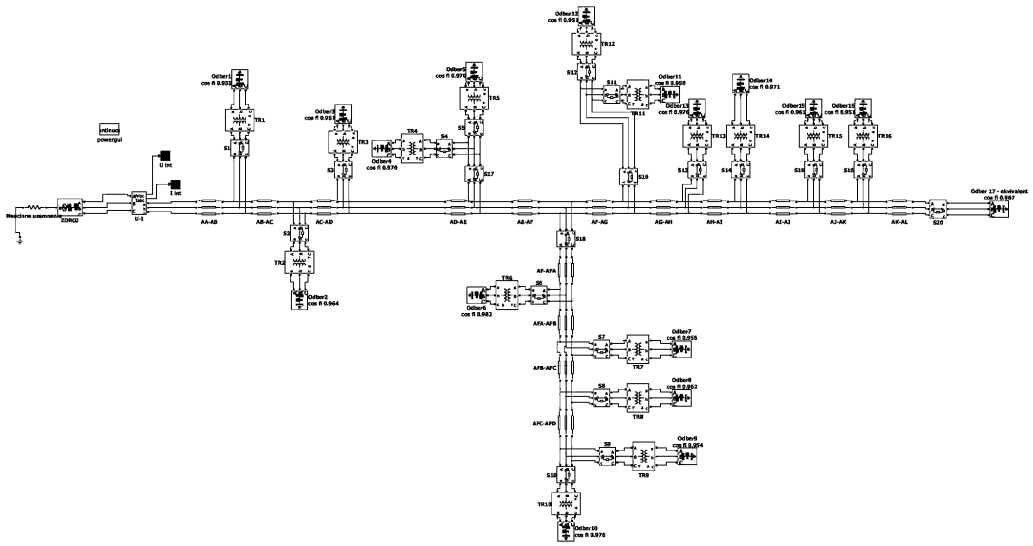
Model siete je určený prostredníctvom parametrov modelov prvkov. Keďže sieť je neúčinne uzemnená, jedným z prechodných dejov môže byť zemné spojenie. Model ES je časť 22 kV siete nachádzajúcej sa na východnom Slovensku. Hodnoty transformátorov, vedení a napájacieho zdroja boli zadané podľa poskytnutých údajov. Hodnoty jednotlivých odberov sú určené na základe optimálneho zaťaženia v prevádzke. Sústava má 17 odberných miest, každé odberné miesto má svoj spínač na odpojenie, resp. pripojenie. Pri viacerých odberoch v uzle je umiestnený spoločný spínač, odpínajúci celú sekciu.

ISBN 978-80-553-0890-6 © 2012 FEI TUKE



Obr. 1 Schéma vybraného úseku ES

Podľa predloženej schémy na obr. 1 sa v prostredí Matlab/SimPowerSystem vytvorí model siete (obr. 2)..

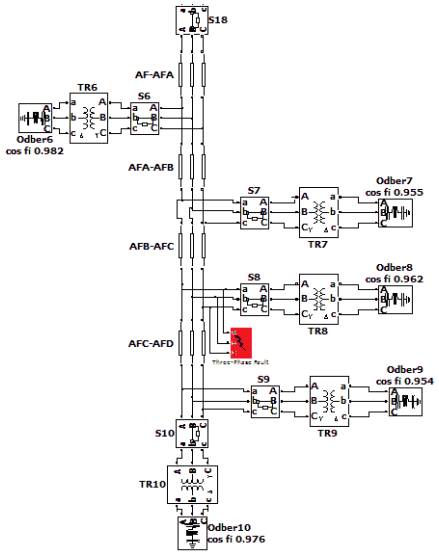


Obr. 2 Schéma vybraného úseku ES v prostredí Matlab/SimPowerSystems

Na obr. 2 je znázornený model ES, jedná sa o neúčinne uzemnenú sieť. V schéme sú umiestnené jednotlivé odbočky, resp. odbery. Daný úsek (sekcia) má dĺžku v hlavnej línii AA-AL 8121 m. Sekcie majú svoje hlavné spínače, ktoré sú označené S17-S20. Tieto spínače, v prípade potreby opravy alebo revízie na odbočke, odpoja celú sekciu od prívodu napätia. Transformátory v odberoch transformujú napätie z 22000 V na 420 V. Hodnoty potrebných veličín, zadané v nastaveniach príslušných modeloch prvkov, boli zadané na základe dostupných technických parametrov.

III. Trojfázový Skrat na Vedení

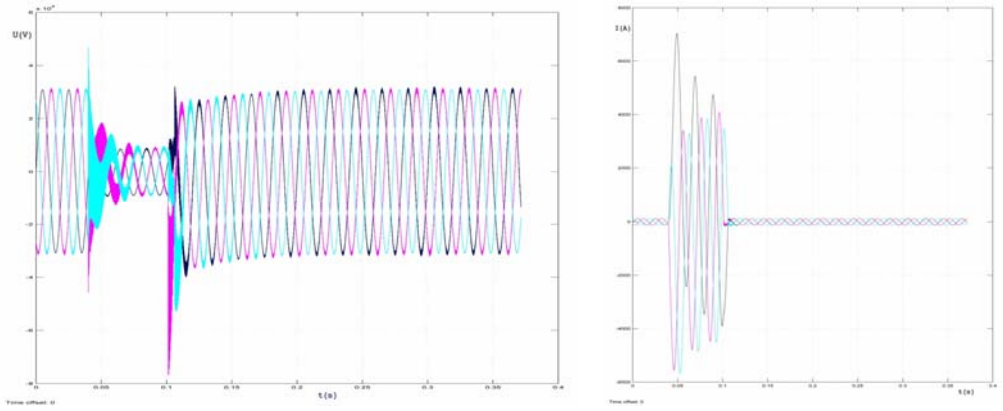
Pri tomto type poruchy dochádza ku kovovému spojeniu všetkých troch fáz. Na obr. 3 je znázornená časť siete, v ktorej došlo k 3-f skratu. V tomto prípade Odber č. 17 je odpojený. Trojfázový skrat nastal pred spínačom Odberu č. 8 v čase $t_1 = 0{,}04$ s a porucha bola odstránená v čase $t_2 = 0{,}1$ s.



Obr. 3 Časť ES so zvýraznením miesta poruchy (3-f skrat)

Ako je vidieť výsledného priebehu na obr. 4a, dochádza k prepätiu vo všetkých troch fázach. Pri vzniku skratu, vo fáze L3 je možné odčítať v čase $t_{k1} = 0{,}0401$ s veľkosť prepätia na úrovni $U_{3k,max} = 32996$ V. Pri odstránení poruchy (v čase 0,1 s) dôjde následne k opätovnému vzniku prepätia s najvyššou hodnotou vo fáze L2 na úrovní $U_{2K,max} = 54327$ V (v čase $t_{k2} = 0{,}1017$ s), v zápornej polperióde. Pre eliminovanie prepätí a rázových (impulzných) prúdov sa preto odporúča, tak ako v tomto prípade, využiť zvodiče prepätí.

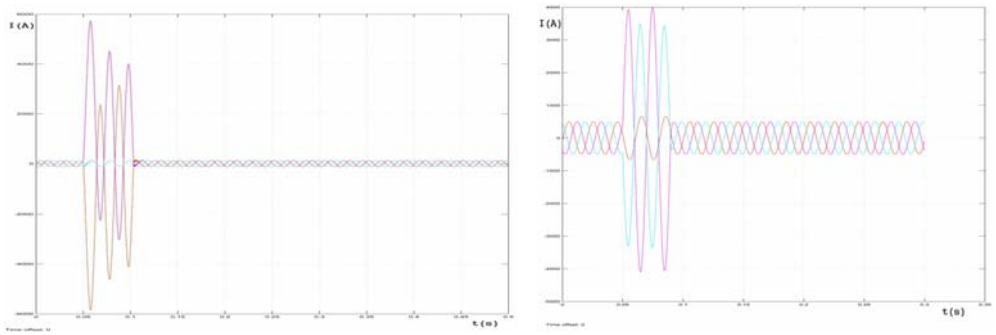
Počas skratu dochádza k vysokému nárastu prúdu z hodnoty $I = 89{,}02$ A na hodnotu špičkového skratového prúdu vo fáze L3 $i_{p3} = 4977$ A (v čase $t_1 = 0{,}049$ s) (obr. 4b). Takéto nárazové hodnoty prúdu môžu byť veľmi nebezpečné pre zariadenia v sieti. Pri dimenzovaní zariadení a nastavovaní nadprúdových ochrán je preto potrebné dbať na správne nastavenie danej ochrany. Pri nesprávnom nastavení ochrán dôjde k dynamickým a tepelným účinkom skratového prúdu, ktoré môžu poškodiť zariadenia v sieti.



a) b)

Obr. 4 Priebeh napätia (a) a prúdu (b) pri trojfázovom skrate

Podobným postupom je možné zistiť veľkosti skratových prúdov pri iných typoch porúch, ako je uvedené napr. na obr. 5.



a) b)

Obr. 5 Priebeh prúdu pri 2-fázovom kovovom (a) a 2-fázovom zemnom skrate (b)

POĎAKOVANIE

Tento príspevok vznikol vďaka podpore v rámci operačného programu Výskum a vývoj pre projekt: Centrum výskumu účinnosti integrácie kombinovaných systémov obnoviteľných zdrojov energií, s kódom ITMS: 26220220064, spolufinancovaný zo zdrojov Európskeho fondu regionálneho rozvoja.

Podporujeme výskumné aktivity na Slovensku/Projekt je spolufinancovaný zo zdrojov EÚ.

LITERATÚRA

[1] NEMERGUT, Lukáš: *Modelovanie elektrizačných sústav využitím nástroja Matlab/SimPowerSystem*. (Modeling of power systems using the tool of Matlab/SimPowerSystem). Diplomová práca. Košice: Technická univerzita v Košiciach, Fakulta elektrotechniky a informatiky, 2012.

[2] KOLCUN, Michal – CHLADNÝ, Vladimír – VARGA, Ladislav – BEŇA, Ľubomír – ILENIN, Stanislav - LEŠČINSKÝ, Peter – MEŠTER, Marián: *Analýza elektrizačnej sústavy*. Košice: FEI-TU, 2005. 419 s. ISBN 80-89057-09-8.

[3] MEDVEĎ, Dušan: *Electric losses modeling of decentralized power sources connection using EMTP ATP*. In: ELEN 2010, ČVUT Praha, 2010, p. 1-9. ISBN 978-80-254-8089-2.

[4] MEDVEĎ, Dušan – HVIZDOŠ, Marek: *Modelovanie v prostredí EMTP – ATP*. 1. vyd. Košice: TU 2011. 74 s. ISBN 978-80-553-0776-3.

[5] MEDVEĎ, Dušan: *Modelovanie prechodných dejov pri pripojovaní rozptýlených zdrojov energie v prostredí EMTP ATP*. In: Elektroenergetika, Vol. 3, No. 7, 2010, p. 15-18. ISSN 1337-6756.

[6] HVIZDOŠ, Marek: *Modelovanie prevádzkových a poruchových stavov v elektrizačnej sústave*. In: E2006/10 – Simulace a dynamické modelování systémů a procesů v elektrizační soustavě, EGÚ Praha, 2006, p. 1-16.

[7] *Documentation – SimPowerSystems*. [Online] Dostupné na internete: < http://www.mathworks.com/help/toolbox/physmod/powersys/ >

Príspevok k bezdrôtovému prenosu elektriny

Matej ORUŽINSKÝ, Juraj KURIMSKÝ

Katedra elektroenergetiky, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

matej.oruzinsky@student.tuke.sk, juraj.kurimsky@tuke.sk

***Abstrakt* — Myšlienka bezdrôtového prenosu elektrickej energie pochádza zo začiatkov 20. storočia. Postupom času vznikali rôzne spôsoby aplikovania tejto myšlienky. Jednou z možností je prenášať elektrickú energiu generovanú Teslovým transformátorom. Tento článok popisuje návrh a realizáciu prenosu elektrickej energie medzi dvoma identickými vinutiami Teslovho transformátora. Sú prezentované merania účinnosti v závislosti na vzdialenosti vysielača a prijímača elektrickej energie.**

***Kľúčové slová* — bezdrôtový prenos, Teslov transformátor, elektrická energia**

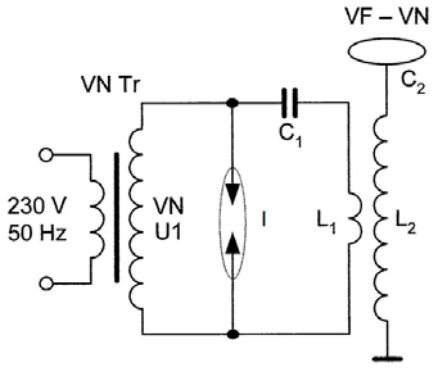
I. Úvod

Teslove zdroje sa tešia veľkej obľube pri experimentovaní s vysokým napätím, ale aj v mnohých oblastiach aplikácie vysokofrekvenčného vysokého napätia. Sú to vysokonapäťové zdroje s transformáciou elektrickej energie na elektro-magnetickú. Vysoké napätie je získavané pomocným vn transformátorom. *„Teslov vn zdroj pozostáva z ladeného primárneho a sekundárneho vinutia transformátora, ktoré pracujú na rovnakom rezonančnom kmitočte. ... Vzhľadom k veľkému rozptylu a koróne majú energetickú účinnosť cca 10 %.*"[1].

V našej práci sme sa venovali návrhu a realizácii Teslovho zdroja ako vysielača elektrickej energie a špeciálnej cievky ako prijímača tejto vyžiarenej energie. Bezdrôtový prenos energie je vyhodnotený meraním účinnosti v závislosti od vzdialenosti vysielača a prijímača.

II. Popis Teslovho transformátora

Primárne vinutie Teslovho transformátora má indukčnosť $L_{1,}$ pozri obr. 1. Je charakteristické tým, že má relatívne počet závitov Pri zapálení iskrišťa sa vytvorí paralelný rezonančný obvod s kondenzátorom $C_{1,}$, ktorý má rezonančnú uhlovú frekvenciu ω_1. Sekundárny obvod $C_2 - L_2$ je ladený na rezonančnú frekvenciu $\omega_{2.}$.



Obr. 1 Schéma zapojenia Teslovho transformátora

Platí vzťah [3]:

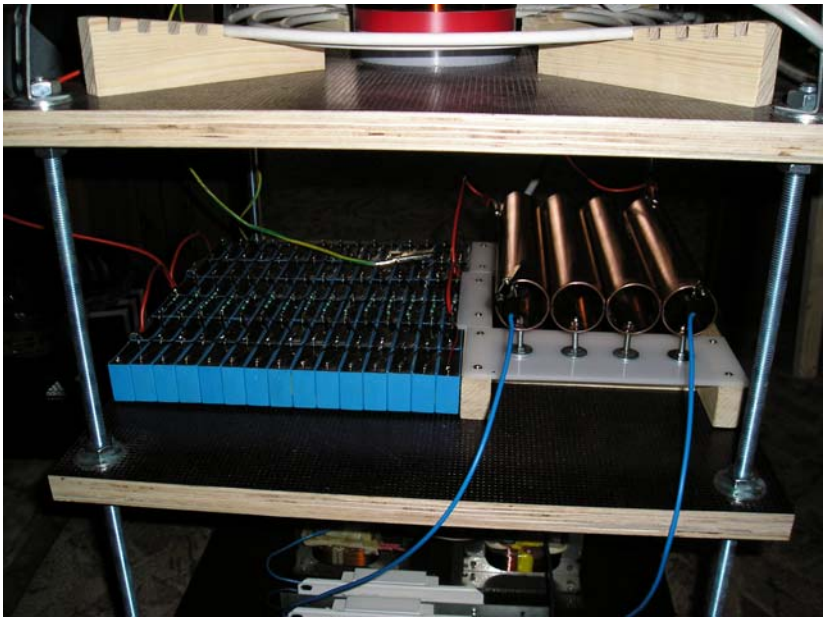
$$\omega_1 = \omega_2 = \sqrt{\frac{1}{L_1.C_1}} = \sqrt{\frac{1}{L_2.C_2}}$$

Vstupný transformátor nabíja cez cievku L_1 kondenzátor C_1 tak dlho, až napätie U_1 dosiahne zapaľovacie napätie iskrišťa. Iskrový oblúk pripojí nabitý kondenzátor C_1 k L_1. V paralelnom obvode vznikne tlmené kmitanie s určitým počtom kmitov, ktoré sa pretransformujú do sekundárneho obvodu [2]. Celý proces sa opakuje.

Vysokonapäťový zdroj -- VN Tr, dodáva napätie, ktoré je väčšie ako zápalné napätie iskrišťa. Ideálne napätie takého zdroja je 4-6 kV, čo zabezpečuje vhodné podmienky pre nastavovanie preskokovej vzdialenosti iskrišťa. Platí, že čím vyššie napätie U_1, tým je potrebný nižší prúd k dosiahnutiu určitého výkonu a z toho vyplývajú nižšie tepelné straty a tiež lepšia celková účinnosť. Naviac, ak je prúd cievkou L_1 veľký, oblúk horí trvalo. Minimálny prúd, potrebný pre správnu činnosť Teslovho transformátora, je daný výškou napätia a časom nabíjania kondenzátora C_1. Pre systém s malým kondenzátorom je potrebný prúd najmenej 15 mA, v praxi je potrebný prúd od 100mA do 1A.

Iskrište, na obr. 1 označené ako *I*, pracuje ako spínač vn. Existuje niekoľko možností pre jeho realizáciu. Najjednoduchšie iskrište je guľové, zmenou vzdialenosti medzi guľovými elektródami nastavujeme veľkosť zápalného napätia. Inou možnosťou je zhotoviť elektródy iskrišťa z medených valcov. Výhodou takejto konštrukcie je rovnomernejšie rozloženie oblúka po ploche elektród, z čoho vyplýva lepšie chladenie a dlhšia životnosť zariadenia. Nevýhodou je problematické nastavovanie zápalného napätia. Poslednou možnosťou je iskrište rotačné, t.j. otáčajúci sa komutátor oproti párom kontaktov. Otáčanie zabraňuje ťahaniu oblúka, zmenou otáčok je možné regulovať výkon. V závislosti od typu pohonu sa rotačné iskrištia delia synchrónne a asynchrónne. Pri synchrónnym iskrišti sú otáčky a uhol natočenia hriadeľa synchrónne s frekvenciou siete. Asynchrónne iskrište, na rozdiel od synchrónneho, používa asynchrónny motor.

Osobitnú pozornosť pri návrhu a dimenzovaní prvkov Teslovho transformátora vyžaduje kondenzátor C_1., zväčša tvorená batériou menších kondenzátorov. Musí znášať veľké zmeny napätia za krátky čas. V tejto práci bol realizovaný sériovo-paralelným spojením kvalitných kondenzátorov do kapacitnej batérie, v odbornej literatúre nazývanej tiež Multi Mini Capacitor – MMC.



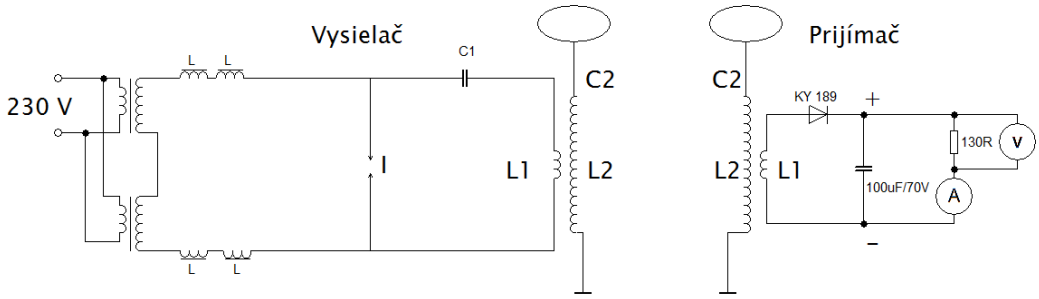
Obr. 1 Kondenzátor MMC a valcové iskrište

Primárne vinutie, tvorené cievkou L_1, má malý počet závitov. Z hľadiska čo najmenších strát sa realizuje z hrubého, najlepšie medeného drôtu. Sekundárne vinutie, cievka L_2 sa vyznačuje

veľkým počtom závitov tenkého, kvalitne izolovaného vodiča, navinutého do tvaru valca. Závity sú vinuté tesne pri sebe, bez vzájomného križovania, pretože by dochádzalo k medzizávitovým prierazom.

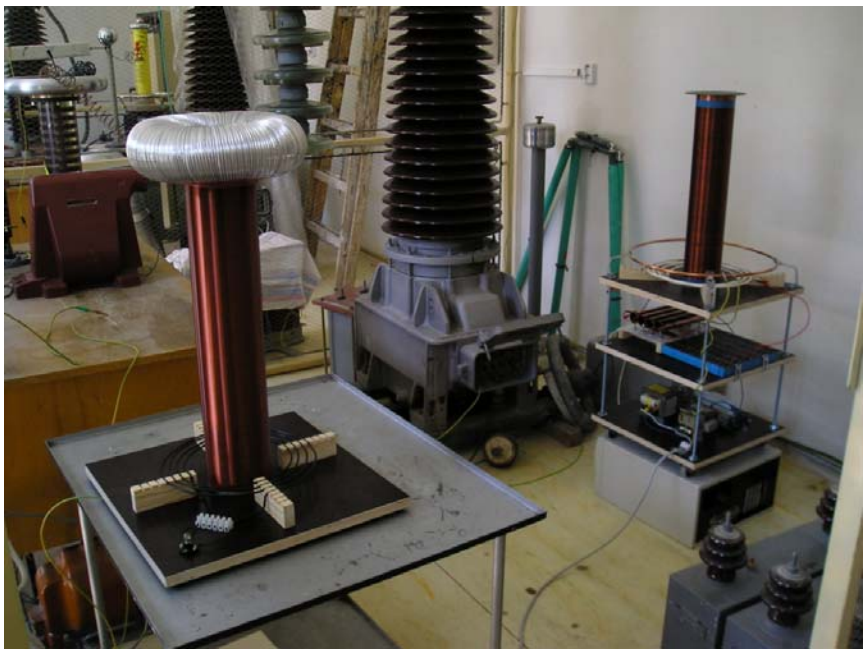
III. BEZDRÓTOVÝ PRENOS ELEKTRINY

Zariadenie na bezdrôtový prenos elektrickej energie sa skladá z dvoch hlavných častí: Prvou je vysielacia časť, ktorá obsahuje vn zdroj, tlmivky (potrebné pre obmedzenie prúdu), iskrište, kondenzátor, primárne a sekundárne vinutie. Druhá, prijímacia časť sa skladá: primárne, sekundárne vinutie, dióda na usmernenie, vysokofrekvenčného prúdu, kondenzátor na odfiltrovanie, záťaž a meracie prístroje.



Obr. 3 Schéma zapojenia bezdrôtového prenosu

Na prijímač je prenesená energia s rovnakou rezonančnou frekvenciou akú má vysielač, čo je f_0=402,589 kHz. Meranie sa uskutočnilo so záťažou s odporom R=130 Ω, merané veličiny boli napätie a prúd so zmenou vzdialenosti. Meranie bolo uskutočnené s dvoma možnými variantmi. Prvou bolo meranie bez použitia toroidu, meralo sa na vzdialenosti od 0,6 do 1,7 metra. Druhé meranie bolo obdobné, naviac sa použil toroid umiestnený na cievku prijímača. So zvyšovaním vzdialenosti klesal aj prenášaný výkon.



Obr. 4 Fotografia prijímača a vysielača

Usmernený signál je vyhladený kondenzátorom 100 μF, ku ktorému je paralelne pripojená záťaž 130 Ω, aby meranie neprebiehalo v stave.

IV. DOSIAHNUTÉ VÝSLEDKY

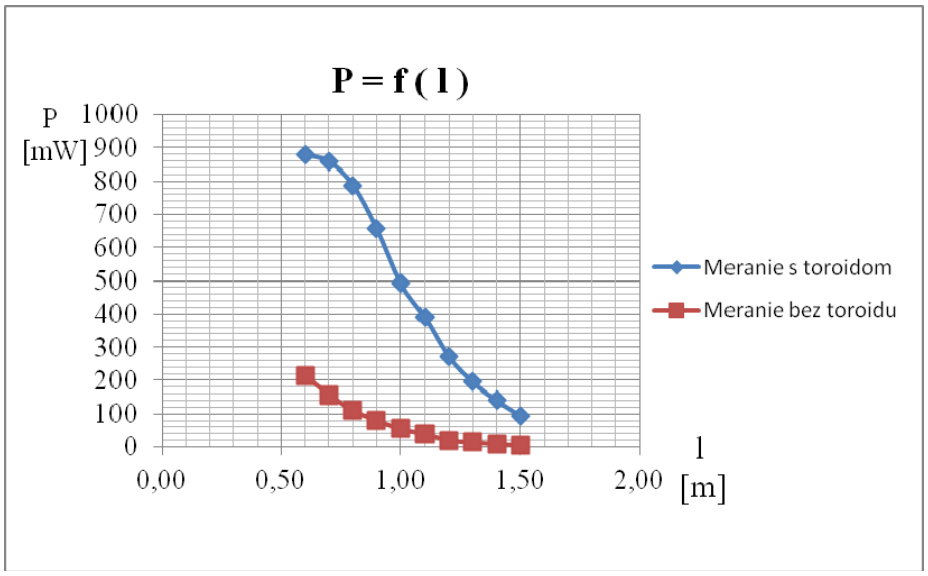
Ako je vidieť v tabuľkách, najvyšší prenášaný výkon bol pri meraní s toroidom vzdialenosťou 0,6 metra. So zvyšovaním vzdialenosti klesal aj prenášaný výkon.

Tabuľka 1
Tabuľka nameraných hodnôt s toroidom

Č.mer.	l [m]	U [V]	I [mA]	P [mW]
1.	1,50	7,65	12,50	95,63
2.	1,40	10,02	14,00	140,28
3.	1,30	12,20	16,00	195,20
4.	1,20	15,30	17,70	270,81
5.	1,10	19,50	20,00	390,00
6.	1,00	23,20	21,20	491,84
7.	0,90	28,50	23,00	655,50
8.	0,80	32,10	24,50	786,45
9.	0,70	35,00	24,50	857,50
10.	0,60	35,10	25,00	877,50

Tabuľka 2
Tabuľka nameraných hodnôt bez toroidu

Č.mer.	l [m]	U [V]	I [mA]	P [mW]
1.	1,50	1,80	3,20	5,76
2.	1,40	2,33	4,00	9,32
3.	1,30	2,90	5,00	14,50
4.	1,20	3,40	6,00	20,40
5.	1,10	4,60	8,50	39,10
6.	1,00	5,65	9,70	54,81
7.	0,90	7,20	11,20	80,64
8.	0,80	8,60	12,80	110,08
9.	0,70	10,80	14,50	156,60
10.	0,60	13,31	16,00	212,96



Obr. 4 Výkon vysielača a prijímača v závislosti od vzájomnej vzdialenosti

V. ZÁVER

Experimentom bolo dokázané, že prenos elektrickej energie pomocou Teslovho transformátora je možný a pomerne jednoduchý. Nastáva však niekoľko problémov, pre ktoré sa tento prenos stal málo efektívny a v praxi nepoužiteľný z hľadiska spotrebovanej elektrickej energie. Najväčším problémom je iskrište, ktoré núti zdroj pracovať do skratu, vďaka čomu spôsobuje vysokú spotrebu elektrickej energie, čo má za dôsledok aj nízku účinnosť. Možnosťou zlepšenia účinnosti by mohla byť vzájomná optimalizácia parametrov prenosových prvkov a vzhľadom na parametre prenášaného elektromagnetického žiarenia.

POĎAKOVANIE

Táto práca vznikla za podpory vedeckej grantovej agentúry ministerstva školstva SR a Slovenskej akadémie vied v rámci projektu č. 1/0487/12.

LITERATÚRA

[1] LYSENKO, Vladimír: VN zdroje. Prvé vydanie. Praha: BEN, 2008. 248s. ISBN 978-80-7300-235-0
[2] KREJČÍRIK, Alexandr: Napájecí zdroje I. Druhé vydanie. Praha: BEN, 1997. 351s. ISBN 80-8656-02-3
[3] LYSENKO, Vladimír: Když se řekne... Zdroje střídavé. In: Automatizace. roč. 50, 2007, č.12, s. 812-813
[4] ORUŽINSKÝ, Matej: Rezonančný zdroj vysokého napätia: Bakalárska práca. Košice: TU FEI, 2011. 52s.
[5] VEVERKA, Antonín. Technika vysokých napětí. Tretie vydanie. Praha: SNTL, 1982. 304s.
[6] NOVÁK, Bohumír: Technika měření vysokých napětí. Prvé vydanie. Praha: Státni nakladatelství technické literatury, 1954. 144s.

Optimal Design of a Fuzzy DC Drive Controller

Pavol FEDOR, Daniela PERDUKOVÁ

Department of Electrical Engineering and Mechatronics, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Slovak Republic

pavol.fedor@tuke.sk, daniela.perdukova@tuke.sk

***Abstract* — The paper deals with the method of design of a fuzzy controller the rules of which are based on generating the optimal input vector using a genetic algorithm. The method is applied in a DC drive with non-linear load. The proposed controller is verified through simulation using the MATLAB software package. Achieved results present a simple applicability of this proposed procedure for a wide class of nonlinear dynamic black-box systems.**

***Keywords* — DC drive, dynamic system, fuzzy controller, fuzzy model, optimal control**

I. INTRODUCTION

Fuzzy controllers of mechatronic systems have been finding wider application in technological practice recently thanks to their simple and robust character [1, 2, 3, 4, 5]. They usually have a PID-type structure and the core of their design consists in specifying the rules and membership functions for the particular system controlled [6]. The application of general „metarules“ is not always suitable, especially where higher order nonlinear systems are concerned [7]. If, however, there exists for the system controlled a suitable input signal time sequence that will achieve the control objective in an optimal manner, it is possible to simply design a fuzzy controller for this system on basis of this sequence (of the optimal input control vector of inputs). This vector can be found, for example, with the help of genetic algorithms.

The paper tests the above mentioned method of fuzzification and rules proposal for a PI-type fuzzy controller and its application in a DC drive with non-linear load. The quality of the fuzzy controller designed using the above mentioned method is verified in the paper by simulations in MATLAB, followed by the presentation of the results achieved.

II. PI-TYPE FUZZY CONTROLLER DESIGN METHOD

A PI-type discrete controller is generally described by the equation:

$$u_k = u_{k-1} + q_0 e_k + q_1 e_{k-1} \tag{1}$$

where u_k, u_{k-1} are values of controller output in the relevant sampling steps, e_k, e_{k-1} are values of the control error, and q_0 ,q_1 are parameters of controller. From these follows its possible structure, shown in Fig. 1 [6].

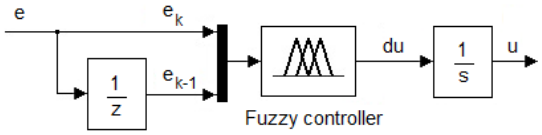


Fig. 1 Block diagram of dynamic system fuzzy model

Its discrete characteristic can then be described by means of the following fuzzy rules:

$$\text{IF } e_k \text{ is ... AND } e_{k-1} \text{ is ... THEN d}u_k \text{ is ..} \tag{2}$$

where the formulations e_k is ... , e_{k-1} is ... and du_k is ... represent the relevant fuzzy areas of the given variables.

The design procedure of the fuzzy controller for the concrete controlled system can be summarized into the following three steps:

1. *Finding the optimal sequence of input values*

The areas and rules for a fuzzy controller can be found in the appropriate database among the appropriate triplets [e_k e_{k-1} du_k] which can be collected from the optimal transition of the controlled system into the desired state. This transition corresponds with a particular sequence of input signal values (hereafter input vector $\mathbf{du}_{opt}$) that best complies with the selected optimality criterion. We chose the optimality criterion in the next form of the integral of the quadratic deviation of the system output *y* from the desired value *w*:

$$J(e) = \int (w(t) - y(t))^2 \, \mathrm{d}t \tag{3}$$

For energy-optimal control of a DC drive the criterion can be chosen in the form of equation:

$$E = \int U_a I_a \, \mathrm{d}t \tag{4}$$

Searching for the $\mathbf{du}_{opt}$ vector requires repeated application of various input vectors to the controlled system (or to its model) and the evaluation of the criterion (3, 4) for each one. The large number of possible input sequences is apparently given by the selected input digitalisation in both the time and value axes and it rapidly rises with the density of this digitalisation. For this reason it is advisable to use the genetic algorithm method in the search for vector $\mathbf{du}_{opt}$, as this method substantially speeds up the process.

2. *Finding the database of optimal data*

If we apply the optimal time sequence of inputs $\mathbf{du}_{opt}$ to the controlled system, we obtain the database of optimal data consisting of triplets [e_k e_{k-1} du_k].

3. *Designing the fuzzy controller from the optimal data database*

From the obtained triplets [e_k e_{k-1} du_k] of optimal data it is possible to design a concrete fuzzy controller of various types using standard procedures of clustering the data into significant clusters and describing them by means of rules.

III. Design of the Torque Fuzzy Controller for a DC Drive

The fuzzy controller design methodology was applied to a DC drive. We will consider a separately excited DC motor with parameters $K_a = 0.625\ \Omega\text{-}1$, $T_a = 0.01$ s, $c\phi = 0.7$ Vs, $J = 0.03$ kgm^2 as a 2nd order dynamic system. If we select state variables of the motor $x_1 = \omega$ (motor angular speed) and $x_2 = I_a$ (motor armature current), then the matrices of its state description will be the following:

$$\mathbf{A} = \begin{bmatrix} 0 & 22.22 \\ -43.75 & -100 \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} 0 \\ 62.5 \end{bmatrix} \tag{5}$$

Considering a fan-type load torque $Mz = k.x_1^2$ ($k = 0.0001$), the simulation structure of the drive and the response of this drive to connection of input voltage $U_a = 30$V is shown in Fig. 2

1. For finding the optimal input sequence $\mathbf{du}_{opt}$ we shall use the diagram shown in Fig. 3

For the selected input sequence vector **du** as shown in Fig. 4a, the time courses of the individual variables are shown in Fig. 4b (system output variable) and Fig. 4c (control deviation).

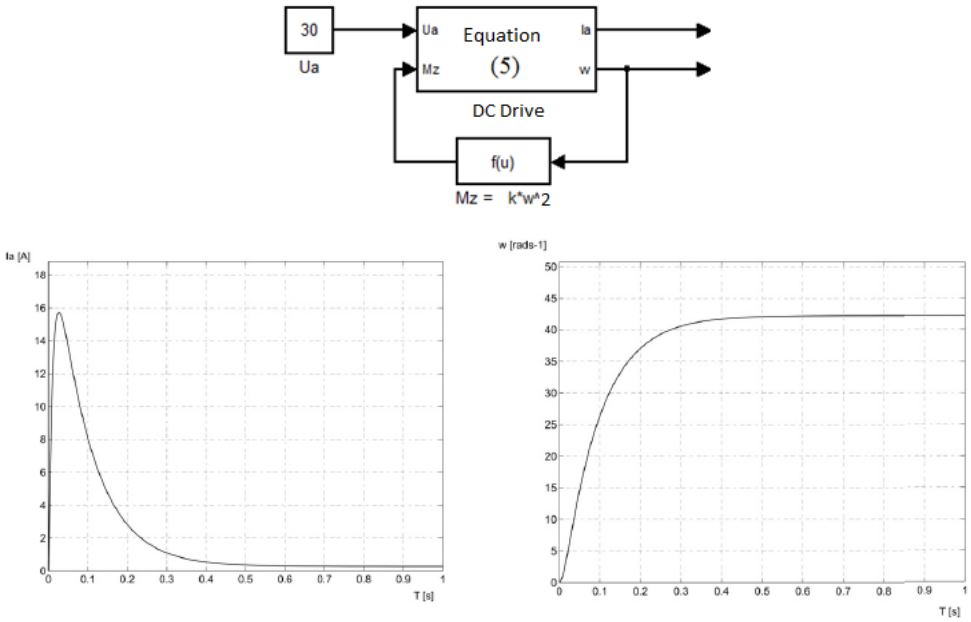


Fig. 2 DC motor response to voltage jump on armature

For the DC drive (5) and desired value w = 50 was found the optimal input sequence vector $d\mathbf{u}_{opt}$ = [200 200 200 150 -20 -10 -4 0 0 -4 0 0 -1 0 0 0 0 0 0 ….] for sampling time T = 0.05 s. The value of the criterion function for the found input vector $d\mathbf{u}_{opt}$ was J = 327.2. The optimal input vector $\mathbf{du}_{opt}$ can be found using the genetic algorithm method. The genetic algorithm was complemented by the direct search method (hybrid function). We used the criterion function J as the suitability measure (eq.3).

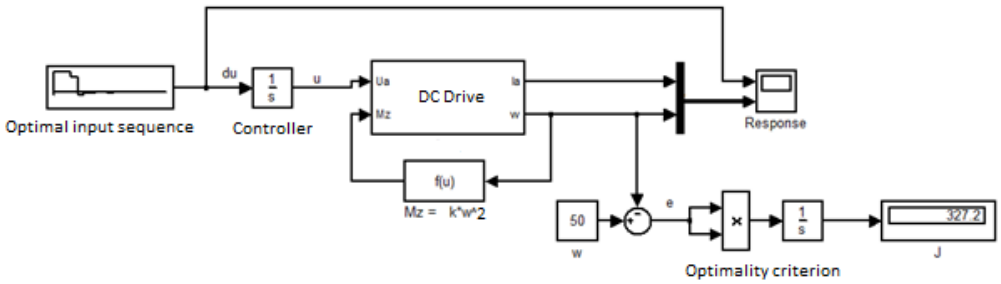


Fig. 3 Block diagram for finding vector $d\mathbf{u}_{opt}$

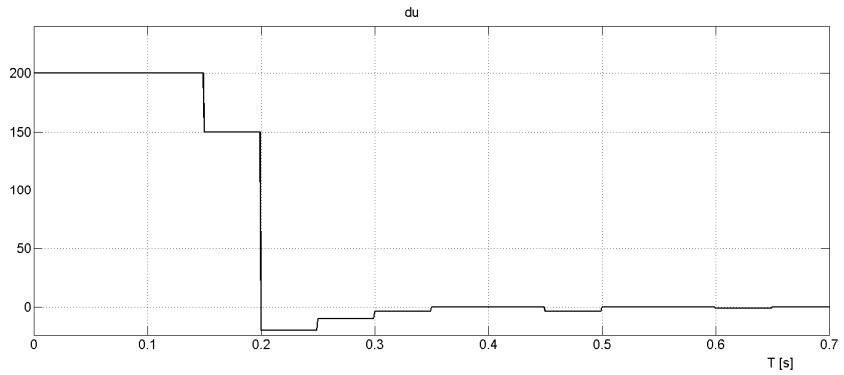


Fig. 4a Time course of variable du

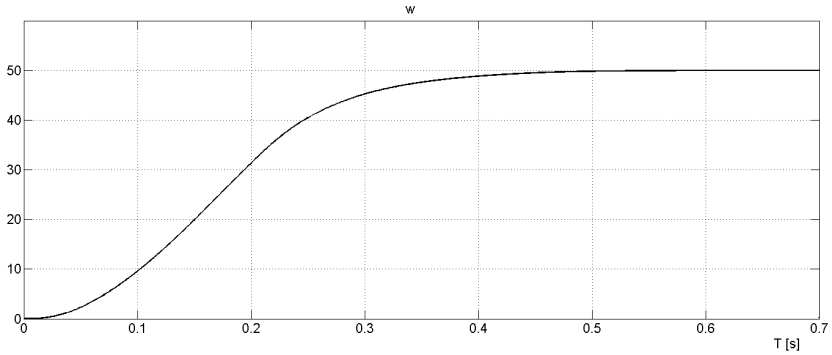


Fig. 4b Time course of system output variable w

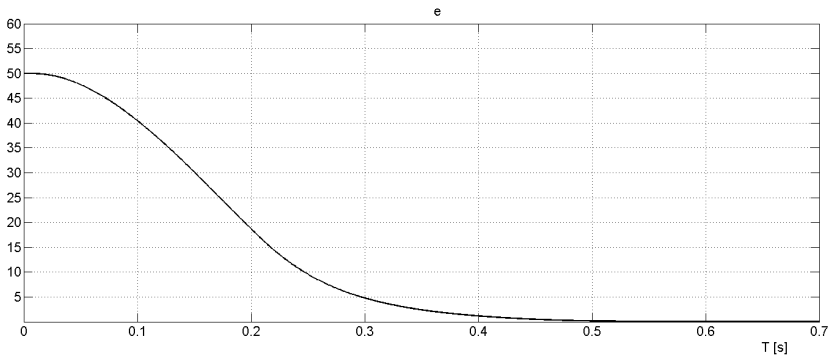


Fig. 4c Time course of control deviation e

2. For finding the database of optimal data consisting of triplets [e_k e_{k-1} du_k] we will use the diagram shown in Fig. 5 where the sampling time T = 0.01 s.

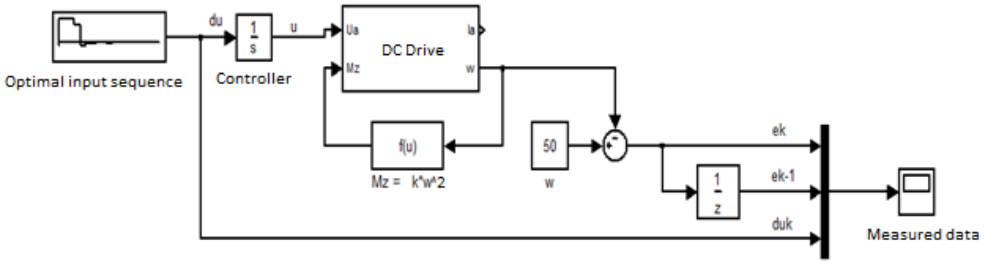


Fig. 5 Measure data for creating the database of optimal data

3. On basis of triplets [e_k e_{k-1} du_k] we will design the FIS structure of a Sugeno type fuzzy controller with nine rules, as shown in Fig. 6.

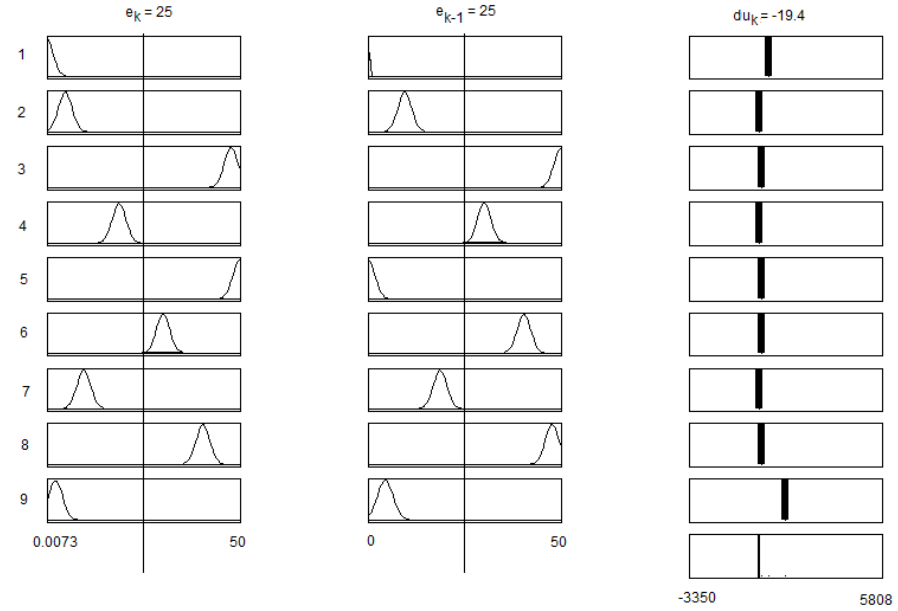


Fig. 6 Internal structure of the designed fuzzy controller

The control circuit using the above designed fuzzy controller is shown in Fig. 7. The quality of control of system (5) output variable w using the fuzzy controller compared to its optimal course w_{opt} for input $\mathbf{du}_{opt}$ is demonstrated in Fig. 8.

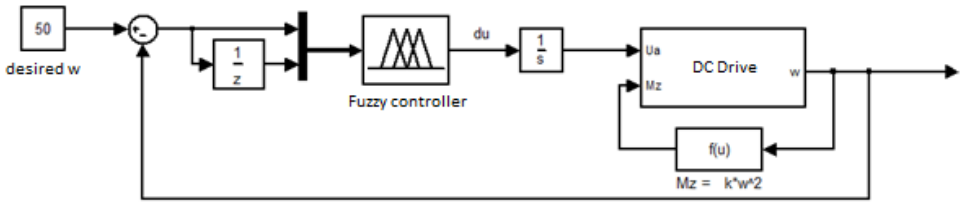


Fig. 7 Control structure with fuzzy controller

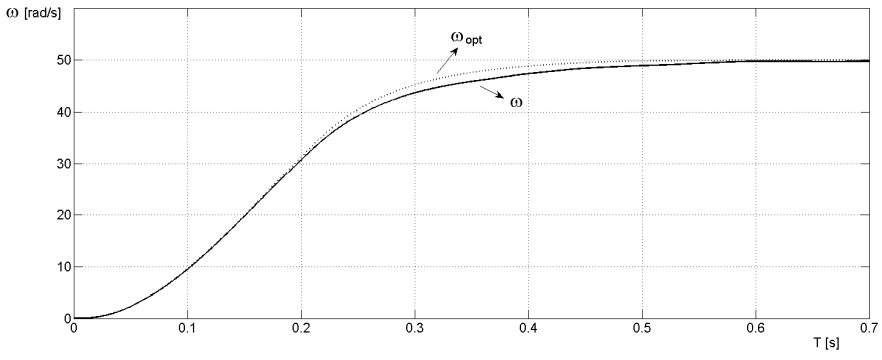


Fig. 8 The dynamics of the optimal and the designed control

Note: For a real controller is need to find the optimal control vectors for the entire workspace and based on them create a database of knowledge for the design of fuzzy controller.

IV. CONCLUSION

The paper briefly describes the design procedure of a PI type fuzzy controller that will establish optimal control dynamics in terms of the selected optimization criterion.

The first step of the proposal procedure is the finding of the optimal sequence of input values through the application of various values of vector **u** to the controlled system input and the evaluation of the selected optimality criterion for each of them. The sampling time for this sequence, i.e. the number of vector **u** samples should be determined in terms of the Shanonn-Kotelnik Theorem so that there are at least 10 of them for the whole transition activity of the closed circuit (e.g. in the example with the DC drive motor which settles in approx. 0.7 s the sampling time of the input sequence was chosen as 50 ms). With the number of samples of the desired vector **u**opt also the complexity of the search for it significantly increases, because the number of input vectors that need to be applied to the controlled system input increases. In case a computer model (at least approximate) of the controlled system is available, the whole process of searching for **u**opt can be fully automated using the genetic algorithm method, as the case was in the examples in this paper. In case we have to deal with a real controlled system, we first either have to generate some type of its computer model from data measured through experiment, or draw on the experience of experts for finding and verifying some suitable sequence of **u** inputs that would become the basis for a fuzzy controller design with similar behavior.

In the second step of the proposal procedure, the input sequence **u**opt is applied to the controlled system and through measurements of suitable variables a database is formed for the fuzzy controller design. The selection of variables for the database and the time of their sampling depend on the specific aims of the control, on the measurability of the controlled system variables, etc. It is possible to measure various different types of databases (ref. the various table forms in the paper) and design various fuzzy controllers for the given system. In all the cases, however, this task can again be fully automated.

The third step of the proposal procedure involves the design of fuzzy controller parameters from the obtained database, keeping in mind that the fuzzy controller is to behave in the same way as the optimal input vector **u**opt and will thus ensure an optimal course of the selected variables of the controlled system. This step involves the use of standard clustering methods and computer tools and can also be fully automated.

The proposed procedure is verified on a concrete example of a DC drive with non-liner load.. Results of simulation experiments show that the controller, in spite of its simplicity and the uncomplicated computer oriented design procedure used, enables considerable improvement of the control circuit dynamic properties also in case of strongly nonlinear higher order controlled systems.

ACKNOWLEDGMENT

We support research activities in Slovakia / Project is co-financed from EU funds. This paper was developed within the Project "Vývoj unikátneho nízkoenergetického statického zdroja pre elektrosystémy ", ITMS 26220220029.

REFERENCES

[1] T. Takagi, M. Sugeno, "Fuzzy identificationof systems and its application to modeling and control," *IEEE Trans. Systems*, *Man and Cybernetics*, 15, 1985, pp. 116-132.
[2] P. Guillemin, "Fuzzy logic applied to motor control," *IEEE Trans. Ind. Appl*., vol. 32, no. 1,1996, pp. 51–56.
[3] W. Li, X. G. Chang, J. Farrell, F. M. Wahl, "Design of an enhanced hybrid fuzzy P+ID controller for a mechanical manipulator," *IEEE Trans. Syst., Man, Cybern., B, Cybern*., vol. 31, no. 6, 2001, pp. 938–945.
[4] F. Barrero, A. Gonzalez, A. Torralba, E. Galvan, L. G. Franquelo, "Speed control of induction motors using a novel fuzzy sliding-mode structure," *IEEE Trans. Fuzzy Syst*., vol. 10, no. 3, 2002, pp. 375–383.
[5] A. Tsourdos, J. T. Economou, A. B. White, P. C. K. Luk, "Control design for a mobile robot: A fuzzy LPV approach," In *Proc. IEEE Conf. Control Applications*, Istanbul, Turkey, 2003, pp. 552–557.
[6] M. Žalman, J. Jovankovič, *Intelligent servosystems*. Bratislava: STU, s. 159, 2007 (in Slovak)
[7] R. Babuska, *Fuzzy Modeling for Control*. Kluwer Academic Publishers, Boston ,1998.
[8] P. Brandstetter, *A.C. Controlled Drives - Modern Control Methods*. VSB-Technical University of Ostrava, 1999.(in Czech)
[9] J. Vittek, S. J. Dodds, *Forced Dynamics Control of Electric Drives*. Žilina: Žilinská univerzita v Žiline, 2003.
[10] Y. S. Zhou, L. Y. Lai, "Optimal Design for Fuzzy Controllers by Genetic Algorithms," *IEEE Trans. On Industrial Applications*, vol.36, no.1, 2000.

Neural Estimators for Shaft Sensorless FOC of Induction Motor

Peter GIROVSKÝ, Jaroslav TIMKO, Jaroslava ŽILKOVÁ

Department of Electrical Engineering and Mechatronics, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Slovak Republic

peter.girovsky@tuke.sk, jaroslav.timko@tuke.sk, jaroslava.zilkova@tuke.sk

***Abstract* — The paper deals with a problem of speed estimation in a shaft sensorless field oriented control structure with induction motor that is based on neural modelling approach. Two different neural estimators were developed; one for observing the magnetic flux and the other one for observing motor angular speed. Structures of the artificial neural network estimators are based on measurable motor variables: components of stator current and voltage in rotating system x-y. Simulation results of the neural angular speed estimator were verified by a Real-Time system. In case of the Real-Time verification the developed neural estimator of rotor angular speed is based on measurable motor variables, components of stator current and voltages in d-q reference system.**

***Keywords* — induction motor, neural network, sensorless control, vector control**

I. INTRODUCTION

Motors play important roles in industrial producing and in many other applications. In their early days, d.c. motors had the advantage of precise speed control when utilized for the purpose of accurate driving. However, d.c. motors have the disadvantage of brush erosion, maintenance requirements, environmental effects, complex structures and power limits. On the other hand, induction motors are robust, small in size, low in cost, almost maintenance-free.

Hasse [9] and Blaschke [10] developed a field oriented control theory to simplify the structure of IM speed control used to drive the d.c. motor. In recent years, the field oriented control theory has become more feasible due to progress in the development of electronics techniques and high-speed microprocessors. Nonlinear control problems can often be solved if full state information is available; in the IM case the rotor states are immeasurable and often the angular speed of the rotor is too costly to be monitored.

In most applications, speed sensors are necessary in the speed control loop. On the other hand, there are applications where lower performance is required, cost reduction and high reliability are necessary, or hostile environment does not allow using speed sensors. In these fields, speed sensorless IM control can be usefully applied. Many different solutions for the estimation of states variables or model parameters have been proposed currently, for example estimators utilizing the motor construction properties, estimators based on the drive dynamic model or estimators based on artificial intelligence [7, 8, 13, 14].

Sensorless controllers, depending on adaptive control and observer theory, on optimal observer design by applying Kalman filter theory [11, 12], on sliding mode control [2, 3], and using artificial intelligence methods [1, 4, 5, 6] have been proposed.

At present, requirements on the dynamic precision are not too strict and virtual or soft sensors are alternatively successfully utilized. Estimators based on artificial intelligence are divided into the following groups:

- systems based on the fuzzy logic,
- systems based on neural networks,
- systems based on hybrid systems,
- systems based on evolutionary algorithms (genetic algorithms).

II. THE NEURAL ESTIMATOR SIMULATION DESIGN

The neural modelling can perform estimation of the induction motor angular speed or other nonmeasurable variables on the neural networks base.

Nowadays, in industry there are commonly used field oriented controlled drives based on different solutions and performance. With field-oriented techniques, the decoupling of flux and torque control commands of the IM is guaranteed, and the induction motor can be controlled linearly like a separately excited DC motor. The DC motor like performance can be obtained in preserving a fixed and orthogonal orientation between the field and armature fields in the induction motor by orientation of the stator current with respect to the rotor flux in order to attain independently controlled flux and torque. Using the field oriented control principle, the stator current component i_{d1} is aligned in the direction of the rotor flux vector and the stator current component i_{q1} is aligned in the direction perpendicularly to it. The rotor flux orientation in the squirrel-cage rotor IM cannot be directly measured, but it can be obtained from terminal variables.

After using transformation of coordinates *d*, *q* to the rotating system *x*-*y*, there is considered, the electric torque is proportional to the i_{1y} component and the relation between the rotor flux and i_{1x} component is given by the first order linear transfer function with the time constant $T2 = L2/R2$.

From this fact and for the considered flux control, the stator current and voltage components were chosen as input signals for reconstruction of the induction motor speed. The developed estimators were trained according to selected training patterns from the direct field oriented control of the induction motor. The block diagram of the control scheme is presented in "Fig 1".

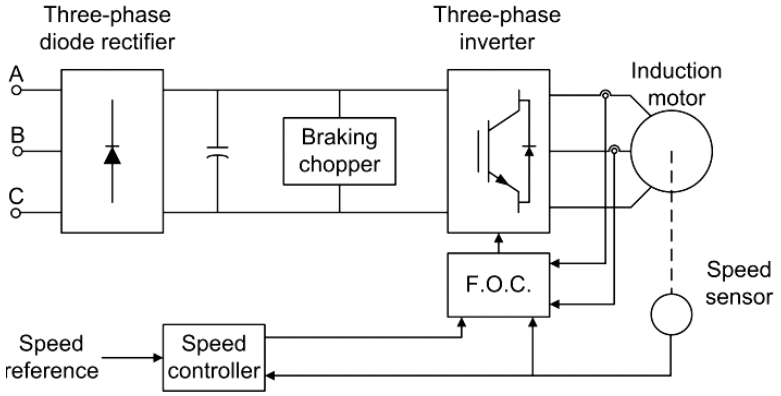


Fig. 1 Basic field oriented control scheme.

A. Simulating design of the magnetising current neural estimator

If for vector control the *x*-th component of the stator current vector is considered as a basis of current-creating component then the magnetising current i_{2m} estimator will process current-creating component of the stator current.

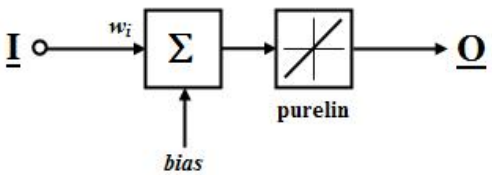


Fig. 2 Basic diagram of magnetising current i_{2m} neural estimator.

As mentioned above, the magnetising current i_{2m} neural estimator bases its estimation on the current-creating component of stator current i_{1x}. Dependence between currents i_{2m} and i_{1x} is linear, and hence the estimator can be made up of a feed-forward neural network without any hidden layer. For the activating function the *purelin* linear function can be used. The input data vector consists of values of the stator current i_{1x} in the step *(k)* and step *(k-1)*, respectively, and also the preceding value of magnetising current i_{2m} in the step *(k-1)*. Basic diagram of such

neural estimator is shown in "Fig. 2".

Here $\underline{O}$ stands for output values vector here, $\underline{I}$ is the input data vector and w_i presents weights of individual connections of neurons. If we have developed the general equation, we shall get the equation for the magnetising current neural estimator in the following form:

$$i_{2m}(k) = purelin\left(\begin{bmatrix} i_{1x}(k) \\ i_{1x}(k-1) \\ i_{2m}(k-1) \end{bmatrix} \underline{w}_i + bias\right) \tag{1}$$

where current $i_{2m}(k)$ is the output variable and the input variables are $i_{1x}(k)$, $i_{1x}(k-1)$ and $i_{2m}(k-1)$.

B. Simulating design of the neural speed estimator

If for the basis of torque-creating component we establish the y-th component of the vector then the speed estimator will estimate this torque creating component from the stator voltage and current.

As it was already mentioned above, the angular speed ω neural estimator bases its estimation on the torque component of stator voltage u_{1y} and current i_{1y}. The relation between the input and output quantities is not represented by a simple linear dependency, and this is reason why for the estimation a cascade neural network with one hidden layer consisted of eight neurons will be used. As an activating function for the hidden layer there was used the *tansig* nonlinear function and for the output layer there was used a purelin linear function. The input data vector is represented by values of stator voltage u_{1y} and stator current i_{1y} in steps *(k)* and *(k-1)*, as well as by value of magnetising current i_{2m} in steps *(k)* and *(k-1)*. Basic diagram of such a neural estimator is shown in "Fig. 3".

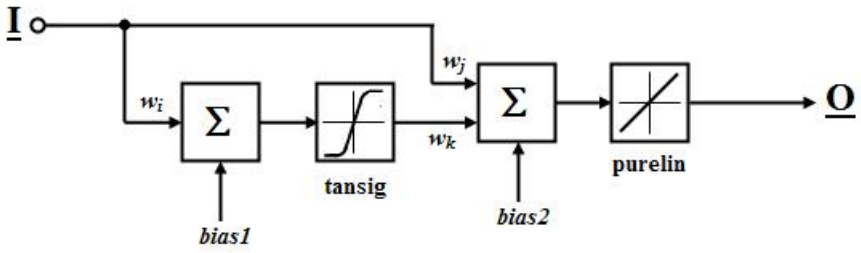


Fig. 3 Basic diagram of ω motor angular speed neural estimator.

In the figure, $\underline{O}$ is the output values vector, $\underline{I}$ present a vector of input variables and w_i, w_j, w_k are weights of individual connections of neurons.

The designed structure of the neural speed estimator can be described by following equation:

$$\omega(k) = purelin\left(\begin{bmatrix} u_{1y}(k) \\ u_{1y}(k-1) \\ i_{1y}(k) \\ i_{1y}(k-1) \\ i_{2m}(k) \\ i_{2m}(k-1) \end{bmatrix} \underline{w}_j + tansig\left(\begin{bmatrix} u_{1y}(k) \\ u_{1y}(k-1) \\ i_{1y}(k) \\ i_{1y}(k-1) \\ i_{2m}(k) \\ i_{2m}(k-1) \end{bmatrix} \underline{w}_i + bias1\right) \underline{w}_k + bias2\right) \tag{2}$$

where the output quantity is $\omega(k)$ angular speed value and where the input are values $u_{1y}(k)$, $u_{1y}(k-1)$, $i_{1y}(k)$, $i_{1y}(k-1)$, $i_{2m}(k)$ and $i_{2m}(k-1)$.

C. Simulation results

In the following there are shown simulation results of sensorless vector control of an induction motor when applying neural estimators of the speed and magnetising current, respectively.

Simulation, design and training of neural estimators were performed for the induction motor with the parameters:

P_n=0,75 kW; U_n= 220V/380V; I_n= 3,8 A/2,2 A;

n_n= 1380 rev/min; p=2; s=0,08; $J=5{,}4.10^{-3}$ kgm^2

"Fig. 4" shows a comparison of real and observed values of the magnetizing current and the angular speed, whilst shown by a dashed line there is the required angular speed value during starting, reversing and loading transients.

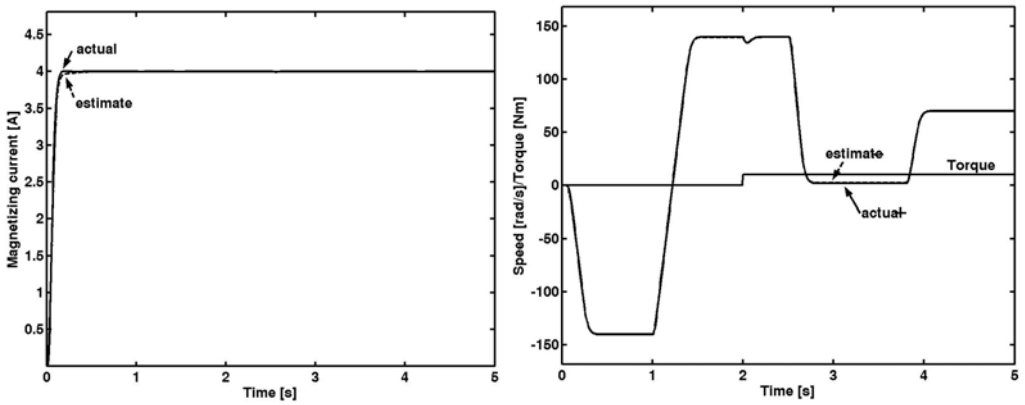


Fig. 4 Comparison of the estimated versus actual: a) magnetising current, b) actual speed of the IM.

The waveforms shown in "Fig. 5" are valid for case of no feedback to control from the neural observers but led directly from the motor mathematical model. In time of 2s the motor was loaded by the rated load torque.

In "Fig. 5" there is shown a simulated response of the induction motor angular speed (by a solid line) at conditions identical with the previous one, shown in "Fig. 4". In this case, and the same as in any following ones, the feedback to control was introduced from neural observers of the magnetising current and angular speed.

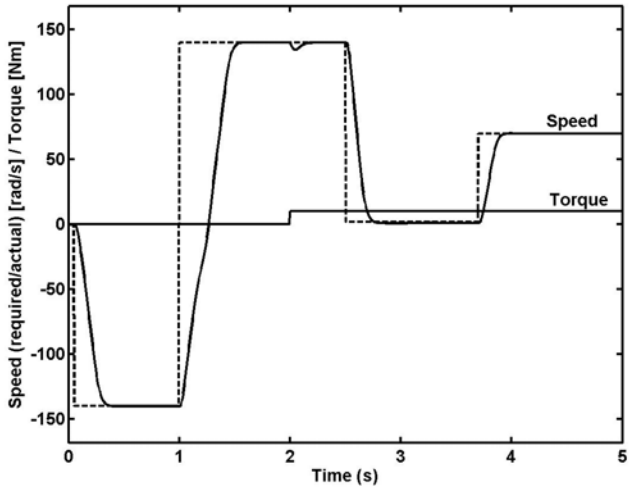


Fig. 5 Transients of desired versus real angular speed and the motor load torque.

III. Experimental Verification

For verification of simulation results an experimental Real-Time system based on RT-LAB system was used. The experimental system consists of the SIMOVERT MASTERDRIVES Vector Control and an induction motor with the same parameters like of the motor used for simulation. As the load there was used a dynamo with resistor and the base of this experimental system consists of Real-time system with NI PCI-6025E.

A. Neural Estimation for Experimental Verification

Regarding different way of vector control in SIMOVERT MASTERDRIVES Vector Control (system in rotary coordinates *d-q*), for design of speed neural estimator was used input stator voltage in step *(k)* – $\boldsymbol{u}(k)$, in the step *(k-1)* - $\boldsymbol{u}(k-1)$ and value of current components *d,q* in step *(k)* – $\boldsymbol{i}_d(k)$, $\boldsymbol{i}_q(k)$ and in step *(k-1)* - $\boldsymbol{i}_d(k-1)$, $\boldsymbol{i}_q(k-1)$. For off-line training using the Levenberg-

Marquardt algorithm was used totally 126013 samples. The output vector for training is represented by value of the rotor speed $\hat{o}(k)$in the step (k).

For the speed neural estimator we used cascade neural network with one hidden layer having six input neurons and six hidden neurons. For the hidden layer activating function there was used the *tansig* nonlinear function and for the output layer we used *purelin* linear function. Using them we received the equation for neural estimator of speed in the following form:

$$\hat{o}(k) = purelin\left(\begin{bmatrix} u(k) \\ u(k-1) \\ i_d(k) \\ i_d(k-1) \\ i_q(k) \\ i_q(k-1) \end{bmatrix} \underline{w}_j + \textbf{\textit{tansig}}\left(\begin{bmatrix} u(k) \\ u(k-1) \\ i_d(k) \\ i_d(k-1) \\ i_q(k) \\ i_q(k-1) \end{bmatrix} \underline{w}_i + bias1\right) \underline{w}_k + bias2\right) \tag{3}$$

B. Experimental results

In the following, the simulation results of sensorless vector control of an induction motor with neural estimators of speed are presented. The principal diagram of vector control with connected neural estimator of speed is shown in "Fig. 6".

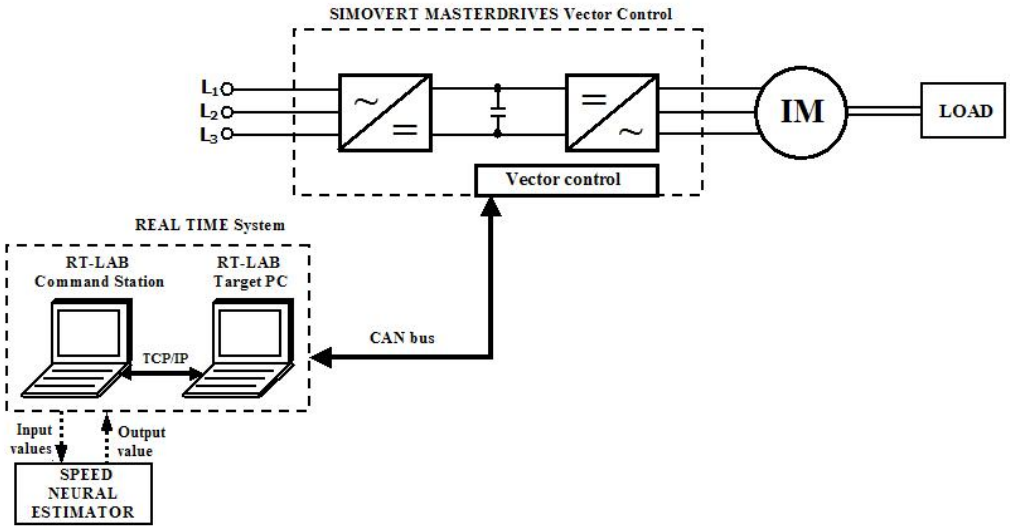


Fig. 6 Principal scheme of Real-Time system with neural estimator.

"Fig. 7" shows a comparison of real value (dash line) and observed value (full line) of the rotor speed. In time of 1 second the required value of rotor speed changed from 0% to 5% of nominal speed; in time 2 seconds from 5% to 50%, in time 3 seconds from 50% to 100% of nominal rotor speed and in time 4 seconds was induction motor reversed.

In the "Fig. 8" there are shown courses of desired (dash line) and actual rotor speed of induction motor in the vector control using the scheme according to "Fig. 6".

In time 1 second the required value of rotor speed changed from 0% to 5% of nominal speed; in time 2 seconds from 5% to 50%; in time 3 seconds from 50% to 100% of nominal rotor speed and in time 4 seconds induction motor reversed.

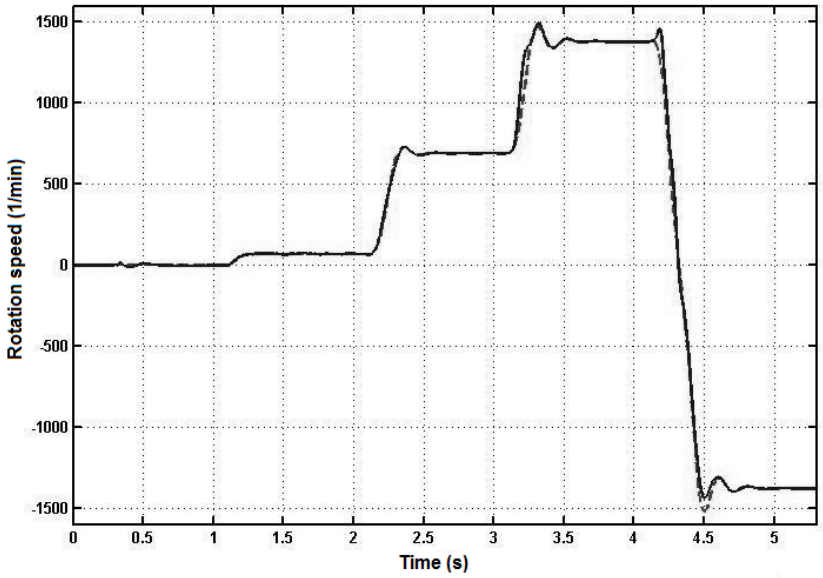


Fig. 7 Comparison of the estimated versus actual speed of the IM.

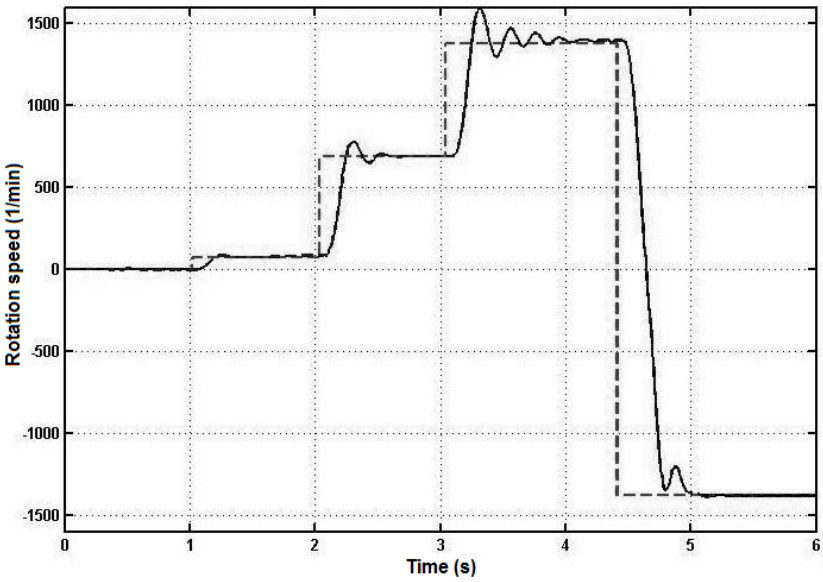


Fig. 8 Time courses of desired versus actual rotor speed.

IV. CONCLUSION

The paper deals with design of the induction motor neural estimators. Based on easily measurable quantities such as stator current and voltage components estimators of the motor speed and magnetising current were designed, whilst utilised were feed-forward and cascade neural networks. Both these networks were trained off-line using the Levenberg-Marquardt algorithm, which is a modification of the traditional back-propagation training algorithm.

At the end of this paper presented are research results. For applied verification used was experimental Real-Time system.

ACKNOWLEDGMENT

The authors wish to thank for the support to the R&D operational program Centre of excellence of power electronics systems and materials for their components II. No. OPVaV-2009/2.1/02-SORO, ITMS 26220120046 funded by European regional development fund (ERDF).

The authors gratefully acknowledge the support of APVV-0185-10 in this work.

REFERENCES

[1] P. Vas, "Artificial-intelligence-based electrical machines and drives", Oxford University Press, Vas P., "Artificial-intelligence-based electrical machines and drives", Oxford University Press, Oxford, 1999.
[2] J. Vittek, S. J. Dodds, P. Makyš, P. Lehocký, "An Observer Design for Forced Dynamics Control of AC Drives", Transcom 2007, Žilina, Slovensko, pp.207-210, 2007.
[3] J. Vittek, P. Bris, M. Štulrajter, M. Pácha, "Chattering free sliding mode control law for position control of the drive employing induction motor", Power Engineering Conference 2008, AUPEC '08, Australasian Universities, pp.1-6, 14-17 Dec., 2008.
[4] M. Kuchar, P. Brandštetter, M. Kaduch, "Sensorless induction motor drive with neural network". IEEE, Annu. Power Elec. Specialists Conf.m pp.3301-3305, 2004.
[5] J. Jovankovič, M. Žalman, "Application of the virtual sensors based on the artificial neural networks", EDPE'03, International conference, Slovakia, 2003, pp.486-490.
[6] Y. Bensalem, W. Abboud, L. Sbita, M. N. Abdelkrim, "A Sensorless Neural Network Speed Control of Induction Motor Drive", Int. Journal of Signal System Control and Engineering Application 1 (2): pp.150-158, 2008.
[7] A. Jadlovská, N. Kabakov, J. Sarnovský, "Predictive Control Design Based on Neural Model of a Non-linear System", Acta Polytechnica Hungarica, Vol. 5, No. 4, pp. 93-108, 2008.
[8] V. Jamuna, S. R. Reddy, "Modeling and speed control of induction motor Drives using neural network", Annals of Diarea de Jos University of Galati, III., Vol. 33, No. 1, pp. 40-49, 2010.
[9] K. Hasse, "Zur Dynamik Drehzahlgeregelter Antriebe mit stromrichtergespeisten Asynchron-kurschlusslaüfer Maschinen", Techn. Hochsch: Darmstadt, Dissertation, 1969, pp.74-78.
[10] F. Blaschke, "The Principle of Field Orientation as Applied to the New Transvektor Closed–Loop Control System for Rotating-Field Machines", Siemens Rev. 39 (5): 1972, pp. 217-220.
[11] S. Meziane, Toufouti R., Benalla H., "Nonlinear Control of Induction Machines Using an Extended Kalman Filter", Acta Polytechnica Hungarica, Vol. 5, No. 4, pp. 41-58, 2008.
[12] J. Žilková, J. Timko, J. Berko, "Speed Sensorless Control of an Induction Motor Drive Using Extended Kalman Filter", In: Acta Technica ČSAV 50, No.4, Prague, 2005, pp. 279-289.
[13] J. Timko, J. Žilková, D. Balara, "Artificial neural networks application in electrical drives", (in Slovak), Calypso s.r.o., Košice, 2002, p. 239, ISBN 80-85723-27-1.
[14] J. Timko, J. Žilková, P. Girovský, "Modeling and control of electrical drives using neural networks", (in Slovak), C-Press, Košice, 2009, p. 202, ISBN 978-80-8086-124-7.

Simulácia meničových systémov v prostredí OrCAD/PSpice a MATLAB/Simulink

Milan LACKO, Peter GIROVSKÝ

Katedra elektrotechniky a mechatroniky, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

milan.lacko@tuke.sk, peter.girovsky@tuke.sk

***Abstrakt* — Článok opisuje možnosti simulácie výkonových obvodov v programoch OrCAD/PSpice ako aj v prostredí MATLAB/Simulink a SimPowerSystems Toolbox. Ďalej sa zaoberá výhodami a nevýhodami oboch programov a ich prepojením. Spojením prostredí MATLAB a PSpice je možné vytvoriť pomerne výkonný nástroj na simuláciu zložitých obvodov, najmä ich riadiacich častí, napríklad pri simulácii spínaných zdrojov.**

***Kľúčové slová* — MATLAB/Simulink, OrCAD/PSpice, PFC preregulátor, SLPS–SimuLink PSpice, SMPS – spínaný zdroj napájania**

I. ÚVOD

Na simuláciu elektrických obvodov v súčasnosti existuje množstvo simulačných programov, v ktorých je možné simulovať činnosť obvodov, od tých najjednoduchších až po tie najzložitejšie. Nespornou výhodou je ich vysoká presnosť, a to aj vďaka veľmi presne popísaným modelom jednotlivých súčiastok použitých pri simulácii. Tieto modely sú neustále dopĺňané, a tak nie je problém simulovať obvod aj s najnovšími súčiastkami, prípadne, ak je potrebné, tieto simulačné programy umožňujú vytvoriť si vlastný model súčiastky. Simulačné programy je možné použiť na simuláciu analógových riadiacich obvodov v spínaných zdrojoch, ako aj samostatných výkonových obvodov. V súčasnosti s nárastom výkonu mikrokontrolérov typu DSC (digital signal controller) sa čoraz viac používajú na riadenie meničov digitálne riadiace obvody. Preto je vhodné už pri simulácii využiť digitálne riadiace obvodu. Avšak pri simulácii zložitejších riadiacich obvodov a najme rôznych regulačných štruktúr v simulačnom programe môže vzniknúť množstvo problémov. Pri veľkej zložitosti simulovanej schémy môže často dochádzať k problémom s konvergenciou výpočtu a taktiež akákoľvek úprava regulačnej štruktúry obvodu môže predstavovať veľký problém. Veľkou nevýhodou je nemožnosť simulácie logických riadiacich obvodov v diskrétnej oblasti.

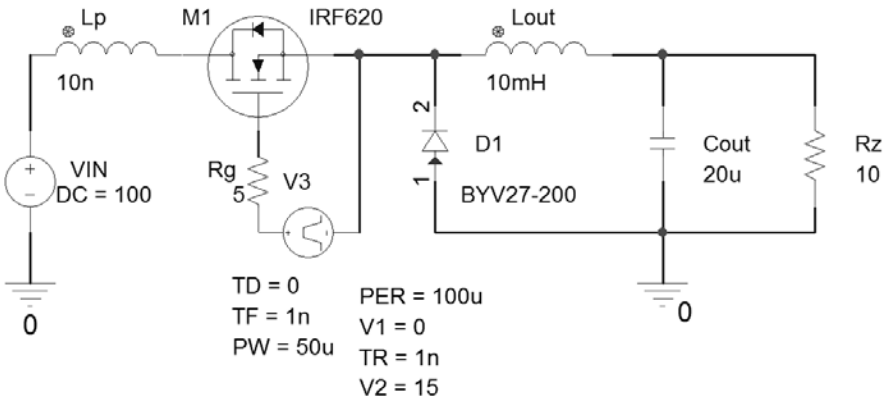
Program MATLAB/Simulink s toolboxom SimPower Systems ponúka široké možnosti pri simulácii elektrických obvodov. Obrovskou výhodou oproti simulácii v prostredí PSpice je možnosť zostavenia zložitých regulačných obvodov využitím takmer všetkých bežne používaných blokov v Simulinku. Aj preto je návrh a ich úprava výrazne zjednodušená. SimPower Systems toolbox využíva pri simulácii veľmi zjednodušené modely prvkov, a z toho vyplýva, že simulácia nedosahuje takú presnosť ako v prípade PSpice. Nie je možné vyšetriť priebeh spínacieho deja výkonového spínača, alebo jeho výkonové straty.

Porovnaním oboch prostredí zistíme, že každý zo systémov ponúka výhody, ktoré predstavujú pre druhý systém problém. Ak tieto dve simulačné prostredia spojíme, a z každého systému využijeme len jeho pozitívne vlastnosti, tak získame výkonný nástroj na simuláciu zložitých obvodov, napríklad práve spínaných zdrojov. Takúto možnosť ponúka simulačný program OrCAD/PSpice od firmy. Cadence. Prepojenie s programom MATLAB zabezpečuje SLPS. V nasledujúcej časti je uvedené porovnanie simulácie jednosmerného znižovacieho meniča (buck converter) v prostredí PSpice a v MATLAB/Simulink.

II. SIMULÁCIA ZNIŽOVACIEHO MENIČA

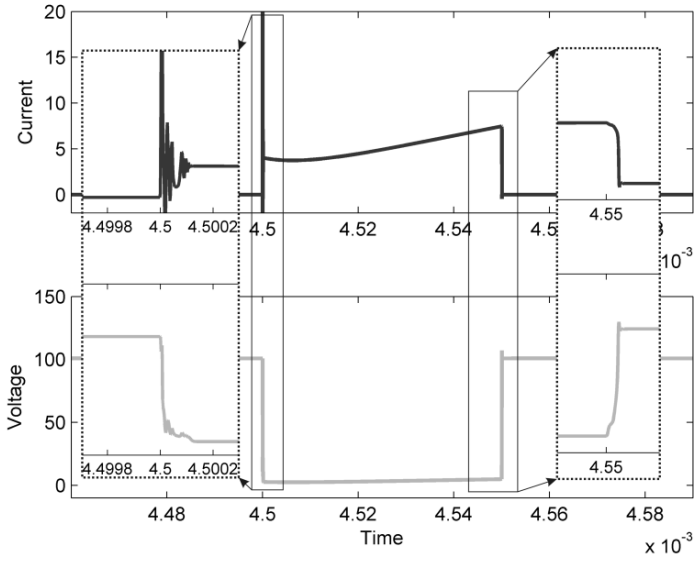
A. Simulácia OrCAD/PSpice

V programe OrCAD/PSpice bola urobená simulácia jednosmerného znižovacieho meniča, ktorého schéma zapojenia je na **Chyba! Nenašiel sa žiaden zdroj odkazov.**.



Obr. 1 Simulačná schéma jednosmerného znižovacieho meniča v programe OrCAD/PSpice.

Menič je napájaný z jednosmerného zdroja s napätím 100V. Napätie je privedené na spínač typu MOSFET, ktorý je spínaní s frekvenciou 10kHz a striedou 50%. Pred spínač bola zaradená indukčnosť L_p, ktorá predstavuje parazitnú indukčnosť prívodov. Nasleduje LC filter tvorený cievkou L_{out} a kondenzátorom C_{out} doplnený nulovou diódou. Na výstup meniča je pripojená odporová záťaž. Na **Chyba! Nenašiel sa žiaden zdroj odkazov.** sú uvedené priebehy napätia a prúdu spínača znižovacieho meniča. V priebehu prúdu sa pri zapnutí spínača objavuje špička prúdu, ktorá je spôsobená pomalým zotavením nulovej diódy. V priebehu napätia sa objavujú prepätia, ktoré sú spôsobené simulovanou parazitnou indukčnosťou prívodných vodičov k spínaču. Výsledky boli prenesené do prostredia MATLAB kvôli zjednoteniu výsledných priebehov.

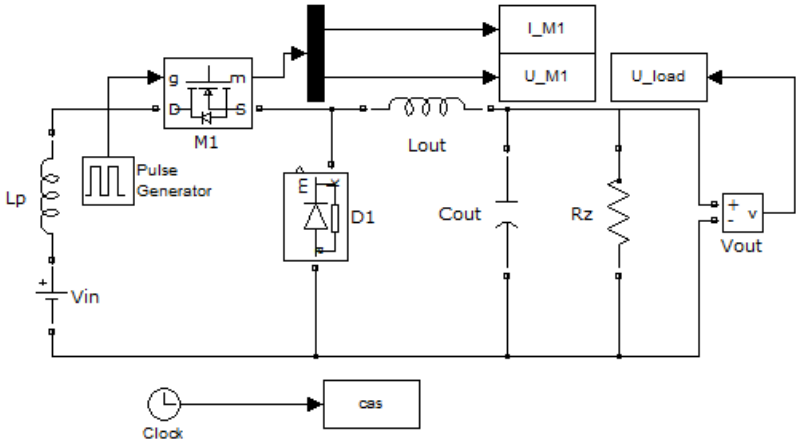


Obr. 2 Simulované priebehy napätia a prúdu spínača v jednosmernom znižovacom meniči v prostredí OrCAD/PSpice

B. Simulácia MATLAB/Simulink

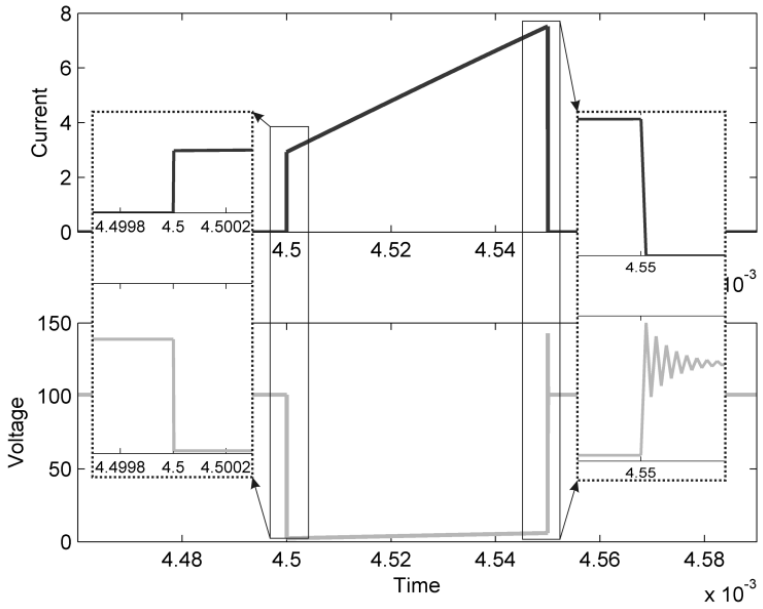
Podobne ako pri simulácii v OrCAD/PSpice bola urobená simulácia znižovacieho meniča aj v prostredí MATLAB/Simulink pomocou toolboxu SimPower Systems. Pomocou blokov bola zostavená simulačná schéma (Obr. 1). Principiálne zapojenie je rovnaké ako v predchádzajúcom prípade. Zásadný rozdiel je v použitých polovodičových súčiastkach. Toolbox SimPower Systems ponúka polovodičové prvky, ktoré majú takmer ideálne vlastnosti.

To spôsobuje pomerne vysokú nepresnosť simulácie, najmä samotných spínacích dejov.



Obr. 1 Simulačná schéma jednosmerného znižovacieho meniča v programe MATLAB/Simulink SimPower Systems

Na Obr. 2 sú uvedené výsledné simulované priebehy v prostredí MATLAB/Simulink. Tu môžeme pozorovať, že priebehy sú takmer ideálne, čo dokazuje aj detail zo zapínania a vypínania výkonového spínača. Táto simulácia dokazuje, že prostredie MATLAB/Simulink nie je vhodné pre simuláciu výkonových obvodov, najmä ak nás zaujímajú rýchle prechodové deje, akým je práve spínanie tranzistora. Avšak, ak chceme vyšetriť teoretické sa správanie obvodu a nepotrebujeme brať do úvahy aj elementárne deje vo vyšetrovanom obvode, tak je MATLAB/Simulink výborný nástroj na simuláciu takýchto obvodov. Ak by sme potrebovali vytvoriť riadenie obvodu, tak by zostavenie regulačnej štruktúry nepredstavoval žiadny problém.

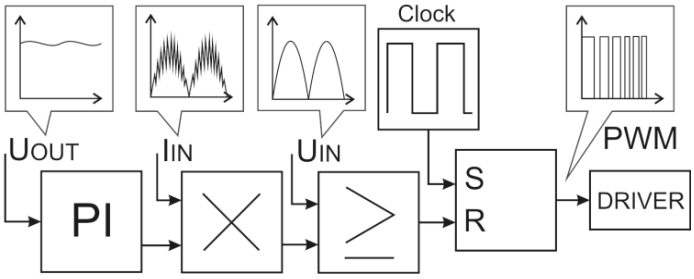


Obr. 2 Simulované priebehy napätia a prúdu spínača v jednosmernom znižovacom meniči v prostredí MATLAB/Simulink

III. Prepojenie prostredí OrCAD/PSpice a MATLAB/Simulink

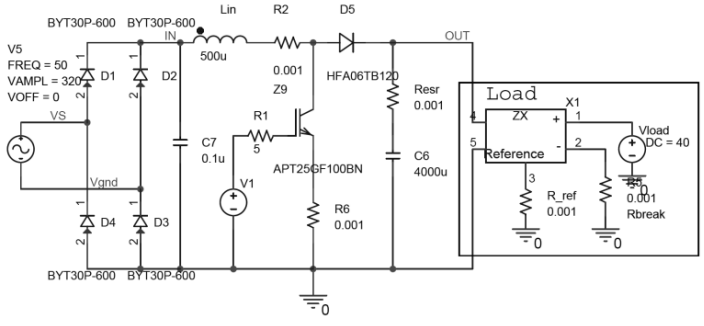
V predchádzajúcej kapitole boli opísané výhody dvoch odlišných simulačných prostredí. Bola urobená simulácia jednoduchého jednosmerného znižovacieho meniča. Obvod bol napájaný konštantným napätím a výkonový spínač bol spínaný s konštantnou striedou a pri konštantnej frekvencii. Takisto sa nemenilo ani zaťaženie meniča. Ak by vznikla požiadavka na zostavenie a simuláciu obvodu, kde by sa vstupné podmienky menili, tak by bolo potrebné doplniť obvod vhodným regulátorom. Ak by sme takýto obvod simulovali v prostredí OrCAD/PSpice, tak môže vzniknúť problém práve s návrhom a s následnou simuláciou

riadiaceho obvodu. Naopak, pri použití programu MATLAB/Simulink by nevznikol najmenší problém so zostavením regulačného obvodu, ktorý by mohol byť aj omnoho zložitejší. Problém by predstavovala práve samotná simulácia výkonového obvodu, ktorá nedosahuje také presné výsledky ako v prípade PSpice. Spojením týchto programov pomocou bloku SLPS v MATLAB/Simulink získame výhody oboch programov, a to vysokú presnosť simulácií a možnosť zostavenia regulačného obvodu. V prípade znižovacieho meniča, kde regulovanou veličinou je výstupné napätie, je potrebné zostaviť regulátor, ktorý bude prakticky meniť striedu spínača podľa žiadaného výstupného napätia. Avšak existujú riadiace obvody, kde je potrebné na reguláciu jednej veličiny využiť aj ďalšie namerané hodnoty v obvode. Príkladom môže byť PFC preregulátor, kde je pre reguláciu výstupného napätia potrebné snímať vstupné aj výstupné napätie a takisto aj vstupný prúd. Na Obr. 3 je uvedená regulačná štruktúra PFC preregulátora. Upravená hodnota výstupného napätia je privedená na do PI regulátora. Výstup z PI regulátora sa privádza do bloku násobenia, kde vstupuje aj hodnota vstupného prúdu resp. prúd cievky. Násobok týchto dvoch vstupov sa privádza do porovnávacieho bloku a porovnáva sa s okamžitou hodnotou vstupného napätia. Výstupná logická hodnota je privedená na resetovací vstup RS klopného obvodu a určuje okamih vypnutia riadeného spínača. Do RS klopného obvodu je privedený ešte hodinový signál, ktorý definuje frekvenciu spínania výkonového spínača.



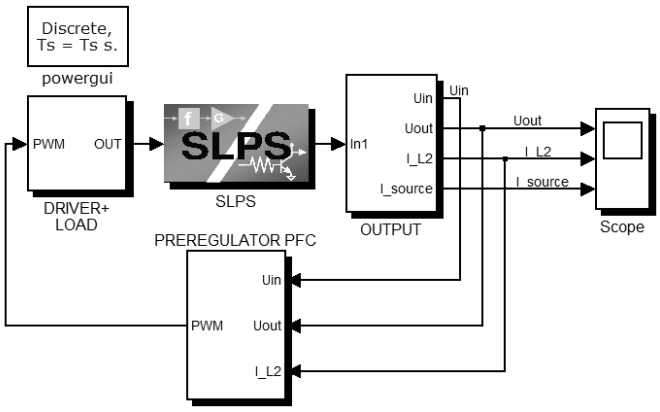
Obr. 3 Regulačná štruktúra PFC preregulátora

V programe OrCAD/PSpice bol zostavený obvod PFC preregulátora, ktorý je znázornený na Obr. 4. Vstupné napájacie napätie s hodnotou 230V a frekvenciou 50Hz je privedené na mostíkový diódový usmerňovač. Za usmerňovačom sa nachádza malý filtračný kondenzátor C_{IN}, ktorého úlohou je filtrovať vzniknuté rušenie spôsobené spínaním výkonového spínača. Nasleduje filtračná cievka L_{IN}, v ktorej sa akumuluje energia v čase zapnutia spínača Q1. Nasleduje usmerňovacia dióda D5. Na vyhladenie výstupného napätia PFC preregulátora slúži výstupný filtračný kondenzátor. Na výstup je pripojená napätím riadená záťaž. Každý napájací zdroj použitý v schéme je možné riadiť pomocou prepojenia SLPS s MATLAB/Simulink a taktiež výstupné veličiny ako prúd, napätie a výkon jednotlivých prvkov obvodu je možné preniesť do prostredia MATLAB/Simulink.



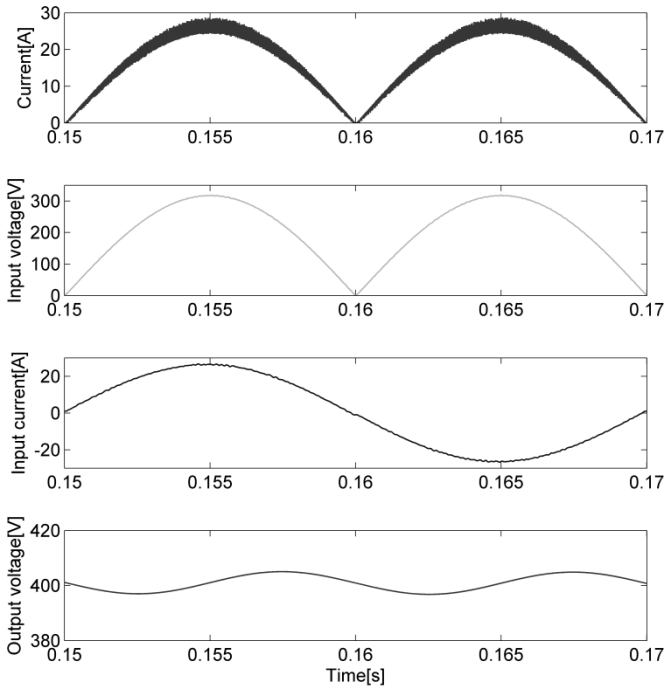
Obr. 4 Simulačná schéma PFC preregulátora

Po zostavení schémy výkonového obvodu bol vytvorený regulačný obvod v prostredí MATLAB/Simulink. Na prepojenie oboch prostredí bol použitý blok SLSP tak, ako je to zobrazené na Obr. 5.



Obr. 5 Prepojenie regulátora s blokom SLPS v prostredí MATLAB/PSpice

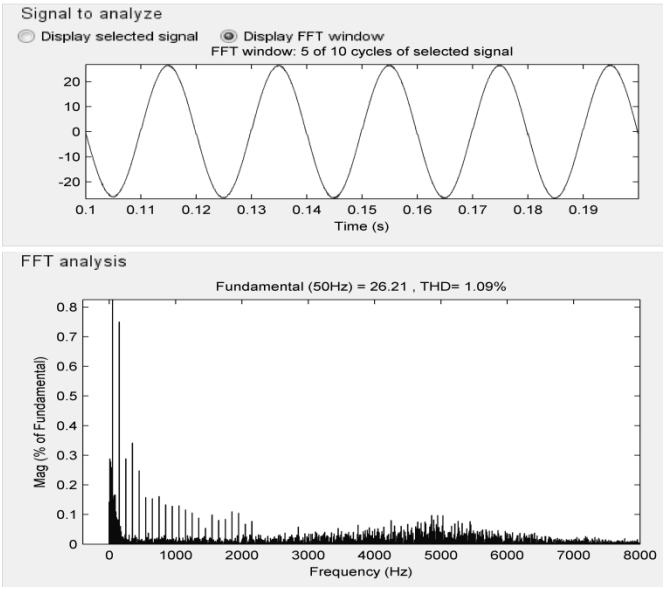
V bloku SLPS je možné zvoliť ovládané zdroje a požadované výstupné signály. Následne bolo zapojenie odsimulované a získané výsledky sú znázornené na Obr. 6. V hornej časti obrázku je priebeh prúdu filtračnej indukčnosti, ktorého stredná hodnota má prakticky sínusový priebeh. Pod priebehom prúdu indukčnosti je priebeh vstupného usmerneného napätia, nasleduje priebeh vstupného prúdu a v spodnej časti je priebeh výstupného napätia. Výstupné napätie má zvlnenie ±5V, čo predstavuje pri napätí 400V zvlnenie 2,5%.



Obr. 6 Simulované priebehy PFC preregulátora, prúd filtračnou indukčnosťou, napätie siete, vstupný prúd a výstupné regulované napätie

Z priebehov na Obr. 6 je evidentné, že odoberaný prúd má praktický sínusový priebeh. Keďže simulácia bola urobená pomocou program MATLAB/Simulink, tak je možné využiť niektoré z ponúkaných funkcií, ako napríklad FFT analýza.

Pomocou Fourierovej analýzy bol skúmaný priebeh vstupného prúdu v čase 0.1 až 0.2 sekundy. V spodnej časti Obr. 7 sú zobrazené jednotlivé zložky vyšších harmonických vstupného prúdu. Základná harmonická (50Hz) má podiel 100% (mimo rozsahu grafu). Ďalej sa v priebehu objavujú aj vyššie nepárne harmonické (3th - 0.74%, 5th - 0.30%, 7th - 0.34%). V oblasti spínacej frekvencie sa takisto objavujú vyššie harmonické. Celkové harmonické skreslenie je na úrovni 1.09% čo predstavuje výbornú hodnotu.



Obr. 7 Analýza vyšších harmonických vstupného prúdu PFC preregulátora

IV. ZÁVER

Porovnaním simulácií obvodov meniča v programoch MATLAB/Simulink a OrCAD/PSpice bolo zistené, že každý je možné využiť na iný typ úlohy. Ak však potrebujeme simulovať detailné správanie sa výkonového obvodu aj so zložitým regulačným obvodom, je vhodné tieto programy prepojiť, čím je možné získať výkonný nástroj na analýzu elektrických obvodov. Taktiež prostredie Matlab/Simulink ponúka množstvo doplnkových funkcií, ktoré je možné využiť pri ďalšej analýze a spracovaní získaných dát zo simulácie.

POĎAKOVANIE

Táto práca bola podporená Kultúrnou a edukačnou grantovou agentúrou Ministerstva školstva, vedy, výskumu a športu Slovenskej republiky pod projektom č.: 042TUKE-4/2012 „Inovácia výučby riadenia mechatronických systémov“.

LITERATÚRA

[1] P. Tobin, “PSpice for Circuit Theory and Electronic Devices“, ISBN: 1598291564, 2007

[2] SimPowerSystems™ 5 User’s guide, available on internet on website: www.mathworks.com/access/helpdesk/help/pdf_doc/physmod/powersys /powersys.pdf

[3] N. Mohan, T. M. Undeland, W. P. Robbins, “Power Electronocs“, ISBN: 0-471-58408-8, 1995, ch. 7

[4] M. H. Rashid, “Power Electronics Handbook“, ISBN: 0125816502, 2001, ch. 19

Využitie ANFIS štruktúry pre odhad rýchlosti AM

Peter NGUYEN, Jaroslava ŽILKOVÁ, Marek VACEK

Katedra elektrotechniky a mechatroniky, Fakulta elektrotechniky a informatiky,
Technická univerzita v Košiciach, Slovenská republika

peter.nguyen@tuke.sk, jaroslava.zilkova@tuke.sk, marek.vacek@tuke.sk

***Abstrakt* — Tento príspevok sa zaoberá možnosťou využitia Adaptívneho neuro-fuzzy inferenčného systému na identifikáciu uhlovej rýchlosti asynchrónneho motora. Navrhnuté boli dva neuro-fuzzy estimátory, kde prvý má vygenerovaný fuzzy inferenčný systém pomocou mriežkového rozkladu a druhý metódou subtraktívneho zhlukovania vstupných údajov. Obidva estimátory boli simulačne overené v tzv. open-loop a close-loop režimoch.**

***Kľúčové slová* — ANFIS, asynchrónny motor, bezsnímačové riadenie, estimátor rýchlosti, vektorové riadenie.**

I. Úvod

V regulovaných pohonoch sa v súčasnosti najviac presadzujú elektrické pohony pozostávajúce z vektorovo riadených asynchrónnych motorov. Asynchrónny motor s kotvou nakrátko (AM) síce z matematického hľadiska predstavuje zložitý nelineárny systém, ale oproti ostatným druhom motorov má veľmi jednoduchú konštrukciu, z čoho vyplývajú aj jeho výhody a to nižšia cena, nízke požiadavky na údržbu, atď.

V dnešnej dobe sa moderné pohony riadia prevažne číslicovo pomocou mikroprocesorov, resp. digitálnych signálových procesorov (DSP), ktoré bez problémov zvládajú realizovať relatívne zložité vektorové riadenie v reálnom čase. Tento dostatočný výpočtový výkon procesorov umožňuje zavedenie ďalších algoritmov (identifikácia parametrov motora, odhad resp. pozorovanie stavových veličín, adaptácia regulátorov, atď.), ktoré slúžia na zlepšenie celkového riadenia regulovaného pohonu. Jednou z možnosti zlepšenia riadenia je aj estimácia resp. pozorovanie rýchlosti motora.

Problematike bezsnímačového riadenia pohonu je v súčasnosti venovaná veľká pozornosť, pretože eliminácia snímača otáčok prináša so sebou niekoľko výhod, ako sú zníženie ceny pohonu (hlavne u pohonoch malých výkonov), zmenšenie rozmeru pohonu, vyššia spoľahlivosť, či menšie požiadavky na údržbu.

Momentálne existuje niekoľko rôznych spôsobov ako možno určiť rýchlosť asynchrónneho motora bez snímača otáčok, pričom jednotlivé metódy možno rozdeliť do troch hlavných skupín:

- estimátory využívajúce konštrukčné vlastnosti motora – vyjadrené póly motora, asymetriu vzduchovej medzery, tretiu harmonickú napätia spôsobenú nasýtením magnetického obvodu, alebo superponovanie signálov
- estimátory a pozorovatele vychádzajúce z matematického modelu AM – priame estimátory využívajúce merané statorové napätia a prúdy, adaptívne systémy s referenčným modelom MRAS (Model Reference Adaptive System) a spätnoväzobné pozorovatele s adaptáciou (Kalmanov filter, Luenbergerov pozorovateľ, pozorovateľ v kĺzavom režime, atď.)
- estimátory a pozorovatele s umelou inteligenciou – fuzzy logika, umelé neurónové siete, genetické algoritmy a ich kombinácia.

Tento príspevok sa zaoberá návrhom estimátora uhlovej rýchlosti založenom na adaptívnom neuro-fuzzy inferenčnom systéme ANFIS (Adaptive Neuro-Fuzzy Inference System) a jeho simulačnom overení v štruktúre vektorového riadenia AM. Navrhnuté sú dva ANFIS estimátory, pričom pre prvý bol vygenerovaný počiatočný fuzzy inferenčný systém pomocou mriežkového rozkladu a v prípade druhého bol vygenerovaný metódou subtraktívneho zhlukovania vstupných dát.

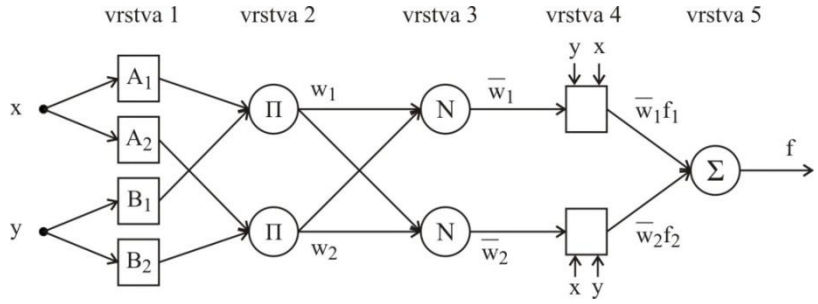
II. ADAPTÍVNY NEURO-FUZZY INFERENČNÝ SYSTÉM

Najčastejšie je rýchlosť motora pri bezsnímačovom riadení identifikovaná pomocou metód vychádzajúcich z matematického modelu AM. Nevýhodou tohto prístupu je potreba mať k dispozícií dobre popísaný systém pomocou matematického modelu a náročnosť návrhu samotného pozorovateľa. V prípade umelej inteligencie (fuzzy logika, neurónové siete, ANFIS) nie je potrebné poznať presný model systému a to predstavuje oproti klasickým metódam jednu z hlavných výhod inteligentných systémov. Taktiež umelá inteligencia je veľmi vhodná či už na reguláciu alebo identifikáciu veličín silne nelineárnych systémov, akými sú aj striedavé motory.

Avšak samotný návrh fuzzy regulátora je náročný, ak je od neho požadovaná vysoká presnosť riadenia. Vo väčšine sa nastavuje metódou pokusu a omylu, čo môže mať za následok nechcené predĺženie návrhu regulátora [1]. V prípade použitia umelých neurónových sieti na aproximáciu rôznych nelineárnych systémov je potrebné mať k dispozícií učiacu množinu dát, ktorá by zahrňovala všetky možné prevádzkové režimy systému. Avšak nie vždy je možné získať takéto dáta.

Štruktúra ANFIS kombinuje výhody fuzzy logiky a umelej neurónovej siete, teda využíva lingvistickú reprezentáciu fuzzy systému s učiacimi schopnosťami neurónovej siete. ANFIS poskytuje niekoľko výhod a to veľmi rýchlu konvergenciu kvôli hybridnému učiacemu algoritmu, schopnosť vytvoriť vhodné vstupné funkcie príslušnosti, kompaktnosť modelu a potrebu menšej učiacej množiny dát [2].

ANFIS predstavuje pätvrstvovú doprednú neurónovú sieť, ktorej výsledkom je fuzzy inferenčný systém typu Takagi-Sugeno [3].



Obr. 1 Štruktúra ANFIS

Prvú vrstvu tvoria adaptívne uzly, ktoré priraďujú vstupným hodnotám stupeň príslušnosti podľa zvolenej funkcie.

$$O_{1,i} = \mu_{A_i}(x), i = 1,2 \tag{1}$$

$$O_{1,i} = \mu_{B_{i-2}}(y), i = 3,4 \tag{2}$$

Kde x (y) sú vstupy i-tého uzla, A_i (B_i) sú označenia lingvistickej premennej a $O_{1,i}$ je stupeň príslušnosti. Teda v tejto vrstve prebieha fuzzifikácia vstupných premenných.

Druhá vrstva je tvorená pevnými uzlami, v ktorých sa násobia prichádzajúce signály.

$$O_{2,i} = w_i = \mu_{A_i}(x)\mu_{B_i}(y), \;\; i = 1,2 \tag{3}$$

Každý výstup z uzla predstavuje váhu w_i pravidla.

Tretiu vrstvu tvoria pevné uzly, ktoré priraďujú pomer váhy i-tého pravidla k súčtu všetkých pravidiel (vykonáva sa tzv. normalizácia váh).

$$O_{3,i} = \overline{w}_i = \frac{w_i}{w_1 + w_2} \tag{4}$$

Vo štvrtej vrstve sa nachádzajú adaptívne uzly, ktorých prenosové funkcie môžu byť konštanty alebo lineárne funkcie.

$$O_{4,i} = w_i f_i = (p_i x + q_i y + r_i) \tag{5}$$

Piata vrstva je tvorená jedným pevným uzlom, v ktorom sa sčítajú všetky do neho vstupujúce signály z predošlej vrstvy, pričom samotný výstup predstavuje výslednú hodnotu ANFIS štruktúry pre dané vstupy.

$$O_{5,i} = \sum_i \overline{w}_i f_i = \frac{\sum_i w_i f_i}{\sum_i w_i} \tag{6}$$

Učenie štruktúry ANFIS prebieha pomocou hybridného učiaceho algoritmu, ktorý kombinuje

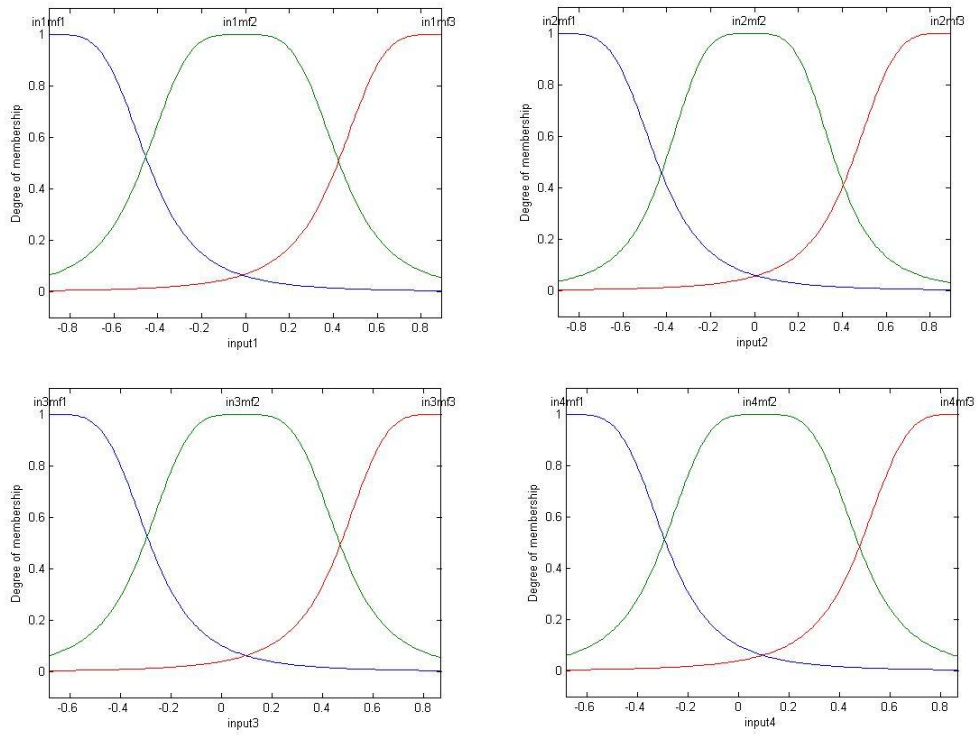
gradientnú metódu a metódu najmenších štvorcov [3]. Každá epocha procedúry hybridného učenia pozostáva z dvoch fáz. V prvej tzv. doprednej sú nelineárne parametre prvej vrstvy nemenné, a pre dané vstupné dáta sa signál šíri cez jednotlivé vrstvy, pričom lineárne parametre v štvrtej vrstve sú identifikované pomocou metódy najmenších štvorcov. V druhej fáze tzv. spätnej sa spätne šíri vypočítaná chyba z výstupu siete na vstup a nelineárne parametre prvej vrstvy sú aktualizované pomocou gradientnej metódy, pričom lineárne parametre ostávajú nezmenené [3].

A. Návrh neuro-fuzzy estimátora uhlovej rýchlosti

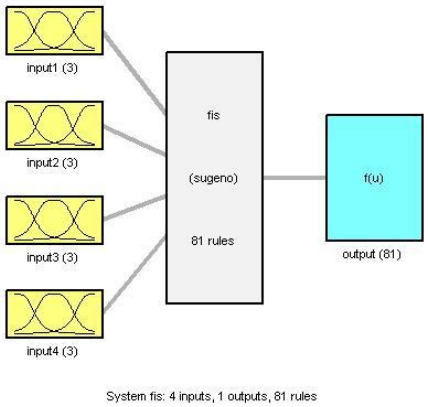
Umelé neurónové siete, fuzzy logika a štruktúra ANFIS predstavujú výborné aproximátory nelineárnych funkcií. V tomto prípade sme sa zamerali na návrh neuro-fuzzy estimátora uhlovej rýchlosti pre potreby bezsnímačového vektorového riadenia AM.

Prvou úlohou pri návrhu estimátora rýchlosti je získanie vstupno-výstupných učiacich dát, na základe ktorých bude učený neuro-fuzzy estimátor. Ďalej je potrebné vybrať vhodné stavové veličiny predstavujúce vstupy estimátora. Po testovaní niekoľkých variantov vstupov boli nakoniec vybrané za vstupy tieto veličiny: u_{1yk}, u_{1yk-1}, i_{1yk} a i_{1yk-1}, teda súčasné a oneskorené y-ové zložky statorového napätia a prúdu.

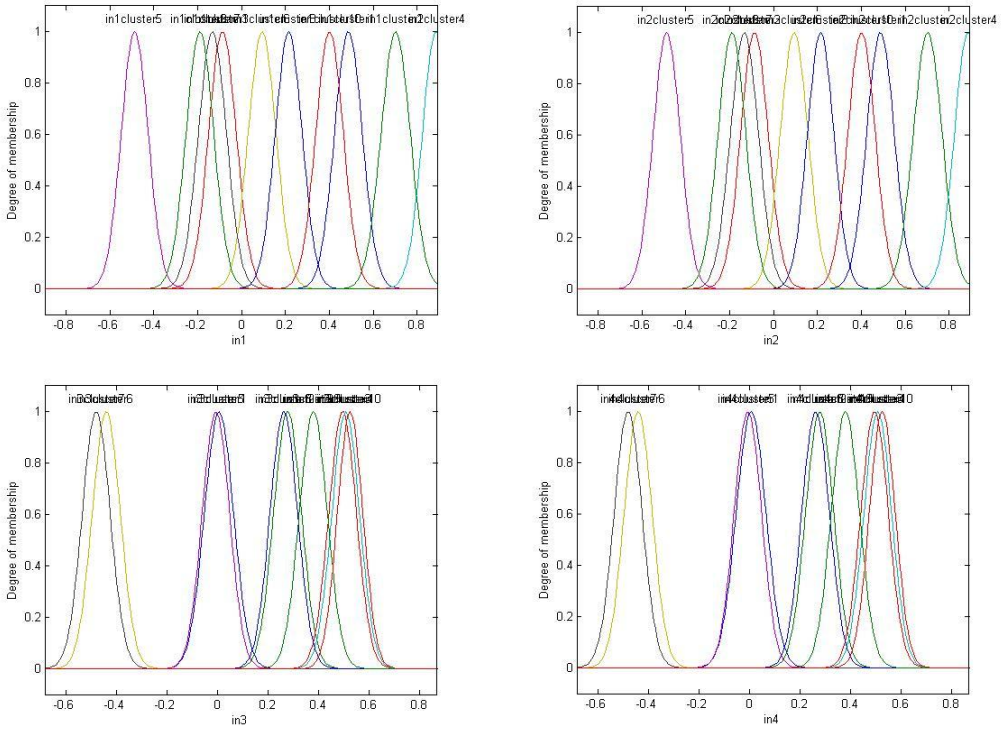
Pre učenie a simuláciu ANFIS estimátora uhlovej rýchlosti bol použitý program MATLAB/Simulink, ktorý obsahuje nástroj na zostavenie a učenie neuro-fuzzy systémov. Zároveň tento nástroj umožňuje generovať východiskový fuzzy inferenčný systém buď na základe mriežkového rozkladu vstupných dát (Grid partition) alebo pomocou metódy subtraktívneho zhlukovania dát (Subtractive clustering). Obidve metódy generovania východiskovej FIS matice sú použité a na obr. 3 a 5 sú zobrazené ich štruktúry.



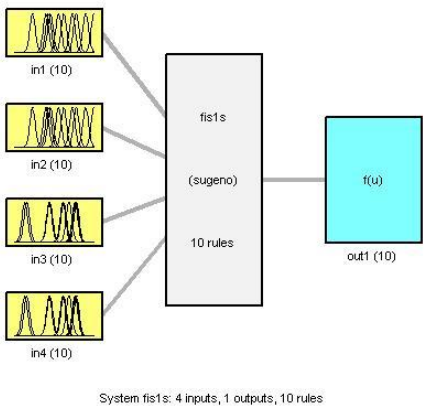
Obr. 2 Rozloženie vstupných funkcií príslušnosti po naučení ANFIS estimátora rýchlosti pre FIS maticu generovanú pomocou mriežkového rozkladu



Obr. 3 Fuzzy inferenčný systém navrhnutý pomocou mriežkové1ho rozkladu vstupných dát



Obr. 4 Rozloženie vstupných funkcií príslušnosti po naučení ANFIS estimátora rýchlosti pre FIS maticu generovanú pomocou subtraktívneho zhlukovania dát



Obr. 5 Fuzzy inferenčný systém navrhnutý pomocou subtraktívneho zhlukovania vstupných dát

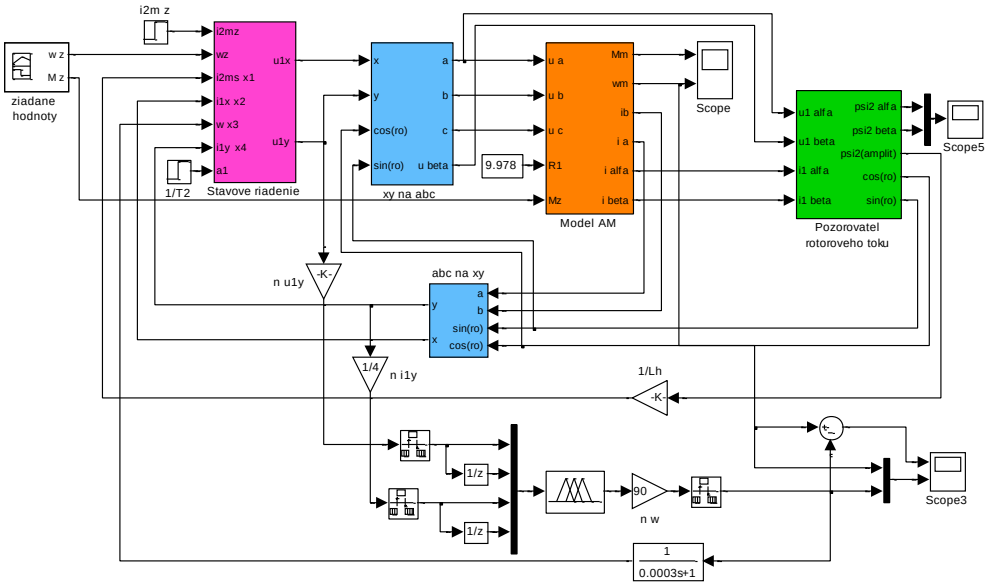
III. SIMULAČNÉ VÝSLEDKY

Obidva navrhnuté ANFIS estimátory uhlovej rýchlosti boli simulačne overené v štruktúre vektorového riadenia asynchrónneho motora s kotvou nakrátko. Parametre AM sú uvedené v tabuľke 1 a celková bloková schéma vektorového riadenia s neuro-fuzzy estimátorom je uvedená na obr. 6.

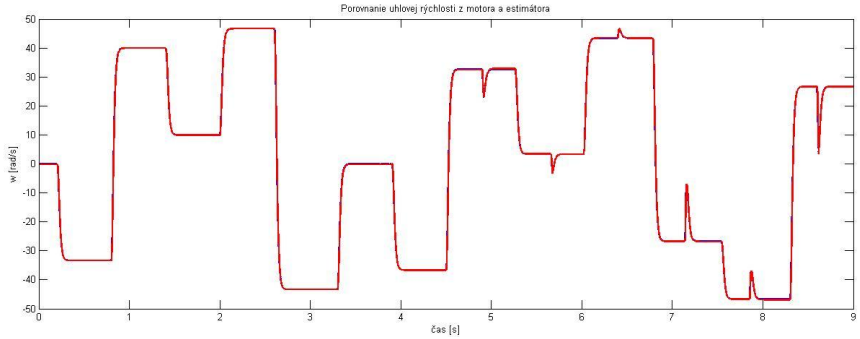
Tabuľka 1
Parametre asynchrónneho motora

Parameter	Hodnota [jednotky]	Parameter	Hodnota [jednotky]
P_n	750 W	J	0,0054 kgm^2
U_n	220/380 V	R_1	9,978 Ω
I_n	3,8/2,2 A	R_2'	9,933 Ω
n_n	1380 rpm	L_h	0,4494 H
p	2	$L_{1\sigma} = L_{2\sigma}$	0,0368 H
s	0,08	$L_1 = L_2$	0,4862 H

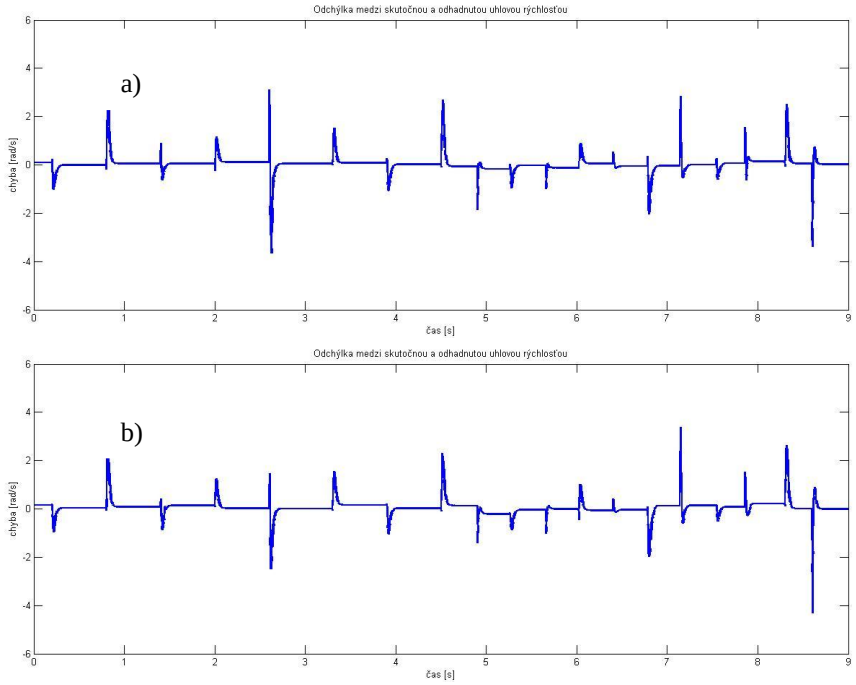
Navrhnuté estimátory boli najprv testované v tzv. open-loop režime, teda výstup z estimátorov nebol privádzaný do riadenia AM, ale iba porovnávaný so skutočnou rýchlosťou motora obr.7. Na obr. 8 sú uvedené chyby odhadu pre jednotlivé estimátory pre rôzne skokové zmeny rýchlosti a záťaže. Ďalej bude platiť označenie ANFIS_1 pre estimátor s fuzzy inferenčným systémom generovaným pomocou mriežkového rozkladu, ANFIS_2 bude označenie pre FIS generované metódou subtraktívneho zhlukovania dát.



Obr. 6 Bloková schéma vektorového riadenia s ANFIS estimátorom uhlovej rýchlosti

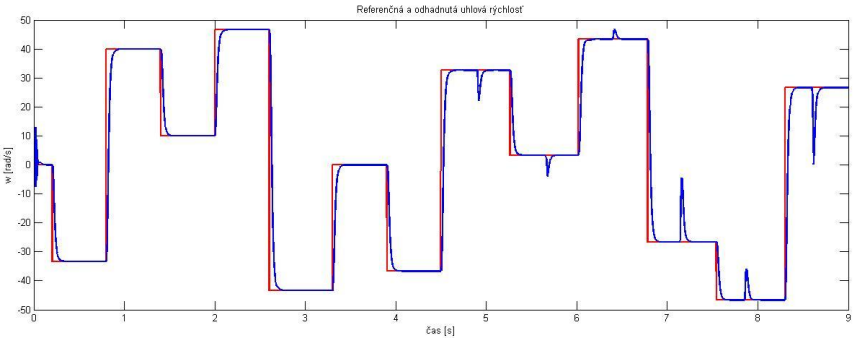


Obr. 7 Porovnanie skutočnej a odhadnutej uhlovej rýchlosti pre ANFIS_1a ANFIS_2 estimátor

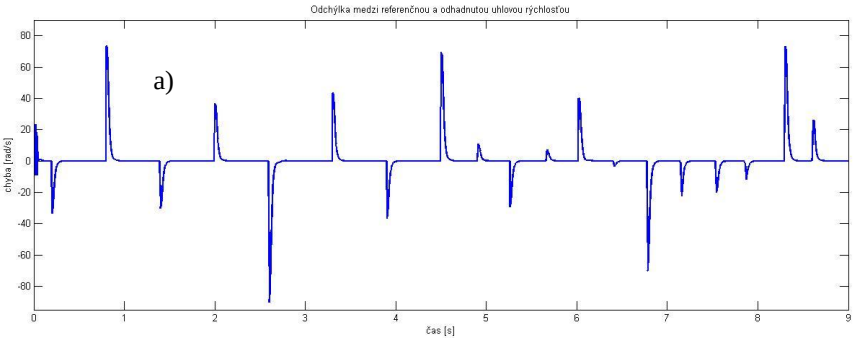


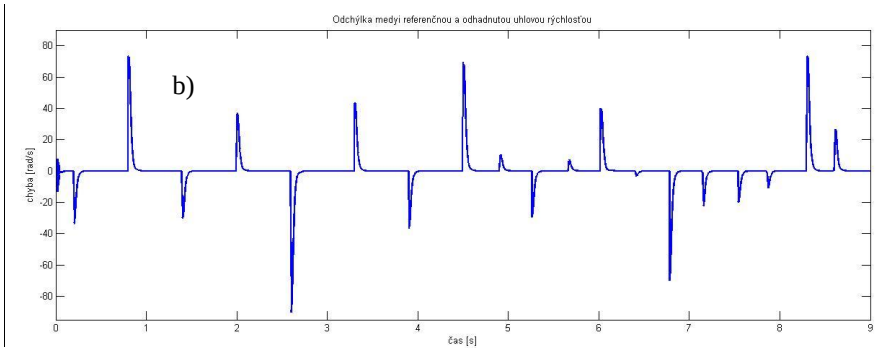
Obr. 7 Chyba odhadu rýchlosti v open-loop režime: a) ANFIS_1 b) ANFIS_2

Ďalšie overenie bolo vykonané pri tých istých želaných hodnotách uhlovej rýchlosti a záťažného momentu, avšak výstup estimátorov bol privedený do riadenia pohonu (tzv. close-loop režim). Obr. 8 zobrazuje želanú a identifikovanú rýchlosť ANFIS estimátora a obr. 9 odchýlku medzi želanou a odhadnutou rýchlosťou.



Obr. 8 Porovnanie referenčnej a odhadnutej uhlovej rýchlosti pre ANFIS_1a ANFIS_2 estimátor (close-loop režim)





Obr. 9 Odchýlka identifikovanej rýchlosti od referenčnej v close-loop režime: a) ANFIS_1 b) ANFIS_2

IV. ZÁVER

V tomto príspevku bola rozoberaná problematika bezsnímačového riadenia AM, kde namiesto snímača otáčok bol použitý ANFIS estimátor uhlovej rýchlosti. Navrhnuté boli dva neuro-fuzzy estimátory. Zo simulačných výsledkov obidvoch estimátorov možno povedať, že kvalita odhadu uhlovej rýchlosti je takmer identická, ale v prípade estimátora s vygenerovaným fuzzy inferenčným systémom pomocou metódy subtraktívneho zhlukovania dát bola získaná ANFIS štruktúra s menším počtom parametrov. Na druhej strane však tieto estimátory vykazovali v ustálených stavoch určité odchýlky (max. odchýlka do 0,2 rad/s), a preto ich možné uplatnenie by mohlo byť v aplikáciách, kde nie je požadovaná veľmi vysoká presnosť riadenia rýchlosti.

LITERATÚRA

[1] R. A. Gupta, "ANFIS Based Intelligent Control of Vector Controlled Induction Motor Drive" 2th International Conference on Emerging Trends in Engineering and Technology, ICETET, pp. 674–680, 2009.
[2] M. Moghadasian, M. Amiri, "Sensorless Speed Control of Induction Motors using Adaptive Neural-Fuzzy Inference System," *International Conference on Advanced Intelligent Mechatronics AIM Hungary*, pp. 1028–1033, 2011.
[3] S. R. Jang, "ANFIS: Adaptive-Network-Based Fuzzy Inference System," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 23, no. 3, pp. 665-685, 1993.
[4] www.mathworks.com – Documentation – Fuzzy Logic Toolbox.
[5] L. Zboray, F. Ďurovský, J. Tomko, Regulované pohony. Vienala, Košice, 2000, pp. 205-225.

Riešenie inverznej kinematickej úlohy sériového robotického ramena pomocou neurónových sietí

Marek VACEK, Jaroslava ŽILKOVÁ, Peter NGUYEN

Katedra elektrotechniky a mechatroniky, Fakulta elektrotechniky a informatiky,
Technická univerzita v Košiciach, Slovenská republika

marek.vacek@tuke.sk, jaroslava.zilkova@tuke.sk, peter.nguyen@tuke.sk

***Abstrakt* — Príspevok sa zameriava na problém určenia polohy jednotlivých kĺbov sériového robotického ramena pri inverznej kinematickej úlohe riešenej pomocou umelých neurónových sietí. Pre riešenie sú použité dva prístupy. V prvom prístupe je navrhnutá jedna neurónová sieť pre určenie všetkých parametrov polohy jednotlivých kĺbov sériového robotického ramena. V druhom prístupe sú navrhnuté viaceré analogické NN pre určenie iba jediného parametra každého kĺbu.**

***Kľúčové slová* — Neurónové siete, inverzná kinematická úloha, sériové robotické rameno,**

I. ÚVOD

A. Neurónové siete

Teória umelých neurónových sietí je založená na poznatkoch z neurofyziológie. Pojem umelé neurónové siete (UNS) bol odvodený na základe analógie s biologickými neurónovými sieťami. Neuróny biologických neurónových sietí sa dlhým vývojom špecializovali na prijímanie, spracovanie, uchovávanie, prenášanie a využívanie informácií.

UNS sa do istej miery pokúšajú napodobniť funkciu ľudského mozgu, ktorý má schopnosť učiť sa, vie paralelne spracovať informácie, pracuje ako asociatívna pamäť závislostí medzi vektorom vstupných a vektorom výstupných premenných.

UNS napodobňujú správanie sa ľudského mozgu hlavne v týchto aspektoch: znalosti sú zbierané v procese učenia, na ukladanie týchto znalostí sú využívané neurónové spojenia, informácie sú spracovávané paralelne

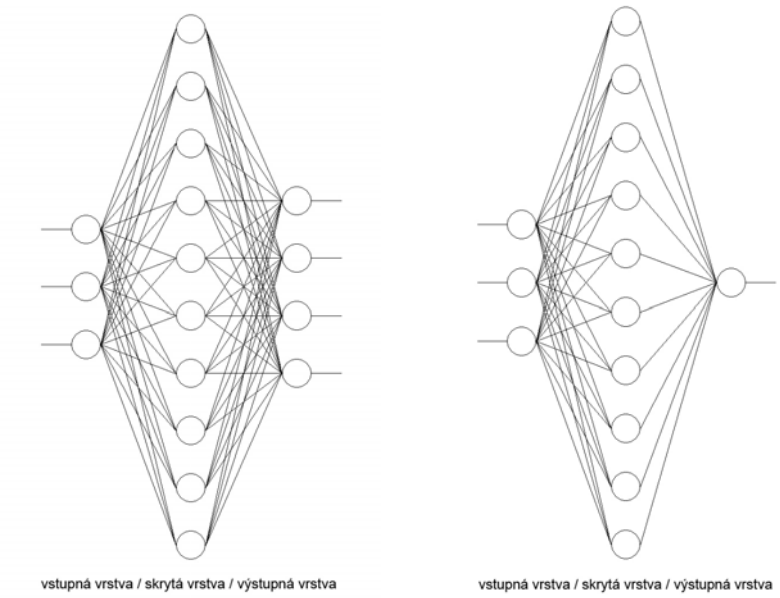
UNS sa významne líšia od iných známych prístupov tým, že majú schopnosť učiť sa, t.j. adaptovať sa podľa vývoja vonkajšieho prostredia k riešeniu danej úlohy, sú univerzálne, keďže dokážu realizovať ľubovoľné zobrazenia, sú robustné, pretože sú schopné abstrakcie, zovšeobecnenia a eliminácie šumu, a lokálne chyby siete majú za následok len postupnú degradáciu funkcie celku, sú veľmi blízke ľudskému chápaniu sveta a vhodné pre prácu s neurčitými, neúplnými a nepresnými informáciami

Zjednodušene môžeme hlavný rozdiel medzi neurónovými a konvenčnými systémami vidieť v tom, že konvenčné systémy pracujú sekvenčne, prevažne podľa určitého presného algoritmu a neurónové systémy pracujú paralelne bez vopred zadaného algoritmu. Ich činnosť je založená na procese trénovania, pri ktorom sa postupne adaptujú k riešeniu danej úlohy.

B. Popis štruktúr vybraných neurónových sietí

Štruktúra neurónovej siete je určená základnými prvkami siete a spôsobom ich vzájomného prepojenia. Pri využití neurónových sietí v modelovaní a riadení systémov je najviac používaným typom dopredná sieť s viacerými skrytými vrstvami, pre ktoré je známe, že trojvrstvové dopredné siete s dostatočným počtom skrytých neurónov dokážu aproximovať s požadovanou presnosťou každé spojité zobrazenie. Na Obr.1 sú uvedené nami použité trojvrstvové dopredné neurónové siete .

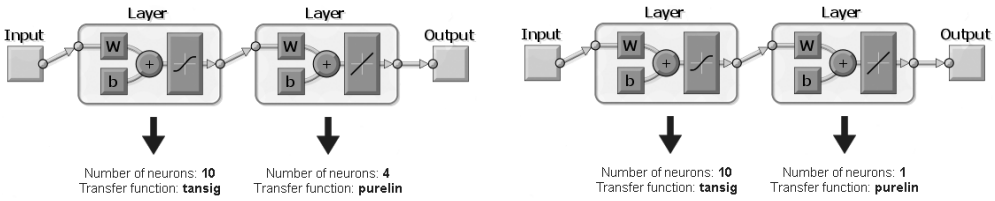
ISBN 978-80-553-0890-6 © 2012 FEI TUKE



Obr. 1 Nami použité dopredné neurónové sieť s jednou skrytou vrstvou

Ak je ako prenosová (aktivačná) funkcia neurónov použitá lineárna funkcia, je dopredná sieť schopná riešiť len lineárne separovateľné úlohy. V prípade, ak sú použité nelineárne prenosové funkcie, bolo dokázané, že neurónové siete tohto typu sú schopné aproximovať ľubovoľnú spojitú funkciu s požadovanou presnosťou, t.j. sú univerzálnym aproximátorom.

Pri riešení úlohy boli v skrytých vrstvách použité tansigmoidné prenosové funkcie a vo výstupných vrstvách sú použité lineárne prenosové funkcie. Takýmto nastavením dokážu skryté vrstvy zabezpečiť nelineárny charakter siete a výstupná vrstva umožní dodržať rozsah výstupu na celom intervale reálnych hodnôt. Pre trénovanie neurónovej siete bolo použité kontrolované učenie Levenberg-Marquardt.



Obr. 2 Schémy neurónových sietí - prvý a druhý prístup

C. Inverzná kinematická úloha

Úlohou kinematickej analýzy je riešenie priamej a inverznej kinematickej úlohy, ktoré sú nevyhnutné pre generovanie riadiacich trajektórii pohybu ramena robota a pre nastavenie polôh jednotlivých kĺbov robota.

Priama kinematická úloha – na základe známych polôh jednotlivých kĺbov hľadá polohu chápadla alebo pracovného nástroja. V prípade paralelného robota je vzájomná poloha určená pomocou dĺžok jednotlivých teleskopických ramien, u sériového robota je rozhodujúcou vzdialenosť a uhly jednotlivých osí.

Inverzná kinematická úloha – na základe známej požadovanej polohy koncového bodu chápadla určuje polohy jednotlivých kĺbov, pomocou ktorých je robot schopný túto požadovanú výslednú polohu dosiahnuť.

Pri inverznej kinematickej úlohe sa vychádza z výsledkov priamej kinematickej úlohy. Bod v priestore, kde sa má dostať koncový bod ramena môže byť popísaný dvoma spôsobmi. Ak požadujeme iba polohu, určíme ju nasledovne:

$$\overline{x_e} = [x \; y \; z]^T \tag{1}$$

kde x, y, z označuje súradnice kartézskeho súradnicového systému. V prípade, že potrebujeme koncový bod ramena dostať do určitého bodu a zároveň pod určitým uhlom, tak pre jeho polohu a orientáciu môžeme napísať:

$$\overline{x_e} = [x \; y \; z \; \psi \; \vartheta \; \varphi]^T$$

kde uhly ψ, ν, φ určujú pootočenie v jednotlivých osiach x, y, z.

Ak určíme polohu bodu v referenčnom súradnicovom systéme, porovnáme tento vektor s výslednou maticou , získanou riešením priamej kinematickej úlohy, ktorá vyjadruje polohu a orientáciu želaného bodu.

$$\overline{x_E} = \overline{B_{0,n}} \begin{bmatrix} \bar{l} & \bar{m} & \bar{n} & \bar{r} \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} l_x & m_x & n_x & r_x \\ l_y & m_y & n_y & r_y \\ l_z & m_z & n_z & r_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (2)$$

Porovnaním Dostaneme sústavu rovníc,

$$\begin{bmatrix} l_x & m_x & n_x & r_x \\ l_y & m_y & n_y & r_y \\ l_z & m_z & n_z & r_z \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \bar{R} & \bar{r} \\ \overline{0^T} & 1 \end{bmatrix} . \overline{p_n} \quad (3)$$

kde $\overline{p_n} = \bar{0}$, ak je posledný súradnicový systém stotožnený s koncovým bodom ramena.

Takto získame dvojnásobný počet rovníc o toľkých neznámych, koľko je kĺbov v danom ramene. Keďže transformačná matica priamej kinematickej úlohy vyjadruje silne nelineárnu závislosť od kĺbových súradníc manipulátora, tak riešením inverznej kinematickej úlohy týmto spôsobom nemusíme dôjsť k výsledku.

Transformačnú maticu môžeme upraviť na tvar:

$$\overline{B_{0,n}} = \overline{T_1} . \overline{T_2} \dots \overline{T_n} \quad (4)$$

Tento vzťah môžeme upraviť prenásobením $\overline{B_{0,n}}$ z pravej aj z ľavej strany:

$$[\overline{T_1} . \overline{T_2} \dots \overline{T_i}]^{-1} . \overline{B_{0,n}} . [\overline{T_{j+1}} \dots \overline{T_n}]^{-1} = [\overline{T_{i+1}} \dots \overline{T_j}] \quad (5)$$

Takto dostaneme 2.n matíc, z ktorých si vyberieme tie najjednoduchšie pre vyjadrenie neznámych q1,q2,...qn.

Po dosadení skutočných hodnôt za q, dostaneme požadované rovnice, z ktorých získame referenčné hodnoty pre učenie neurónovej siete.

II. Simulácie a Merania

Pre simulácie a merania sme sa rozhodli použiť experimentálne robotické rameno so štyrmi stupňami voľnosti a pridaným koncovým nástrojom. Kvôli zovšeobecneniu a možnosti osadenia rôznych typov koncových nástrojov sme sa rozhodli vykonať simulácie a merania bez neho.

Keďže nie sú dostupné informácie o pracovnej obálke tohto robota, určili sme si ju z logického posúdenia a následného overenia na reálnom robote. Nakoľko maximálny dosah robota je 275mm, považujeme ju za hornú hraničnú hodnotu pracovnej obálky. Minimálnu hodnotu dosahu robotického ramena bez koncového nástroja sme stanovili na 60-80mm od základne. Keďže rozsah pohybu každého motora robota je 180°, celkový dosah ramena v horizontálnej rovine je 360°. Ako pracovnú obálku teda môžeme považovať sférický priestor nad základňou robota o výške 275mm s obmedzením súradníc, kam sa robotické rameno nemôže dostať.

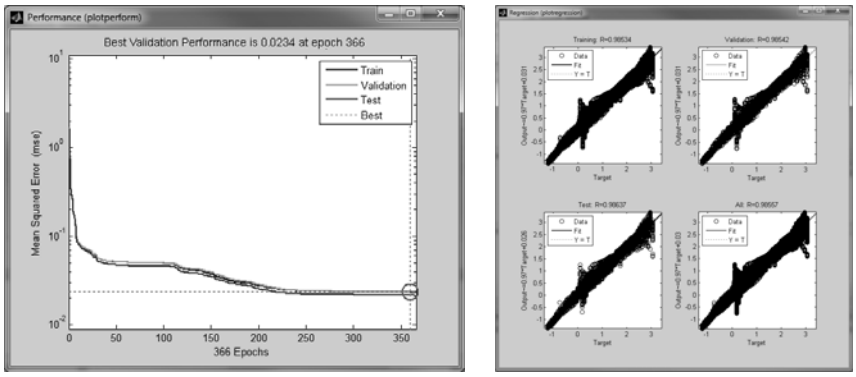
Na určenie vzoriek (súradnice x, y, z) sme sa rozhodli prepočítať zvolenú pracovnú obálku po 5° krokoch v horizontálnom aj vo vertikálnom smere. Výsledkom bolo približne 12 500 vzoriek, z ktorých sme vytriedili tie, ktoré nespĺňajú oblasť pracovnej obálky. Po vytriedení nám zvýšilo 11 298 vzoriek, s ktorými sme ďalej pracovali.

Keďže pre učenie neurónovej siete je nutné mať zadefinovanú databázu vstupov (vzoriek) a k nim príslušnú databázu výstupov, potrebovali sme pre naše robotické rameno prepočet priamej kinematickej úlohy. Neurónovou sieťou následne nahradíme hore uvedený zložitý výpočet inverznej kinematickej úlohy. Pre jej učenie sme využili Neural Networks Toolbox v programe Matlab. Rozhodli sme sa stanoviť si rovnaké parametre pre každú neurónovú sieť a zistiť rozdiel v presnosti a trvaní výpočtu: jednej neurónovej siete pre všetky štyri kĺby robota jednej neurónovej siete pre prepočet iba jedného kĺbu.

Ako typ neurónovej siete sme si zvolili vyššie popísanú doprednú neurónovú sieť s jednou skrytou vrstvou. V skrytej vrstve sme kvôli pomeru rýchlosti a presnosti učenia definovali desať neurónov.

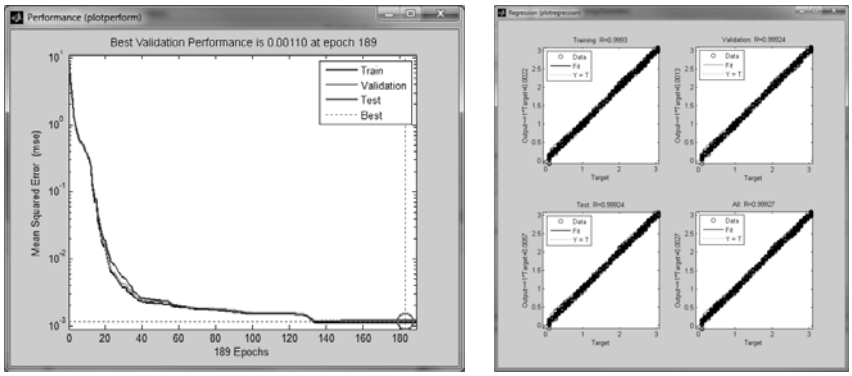
Ako prvú sme učili jedinú neurónovú sieť, ktorá bude prepočítavať všetky štyri stupne voľnosti. V zadaní siete sme zvolili tri neuróny na vstup (vzorky - súradnice x,y,z) a štyri neuróny (štyri kĺby – stupne voľnosti) na výstup neurónovej siete. Po učení siete sme dosiahli

výsledky zobrazené na Obr.3.

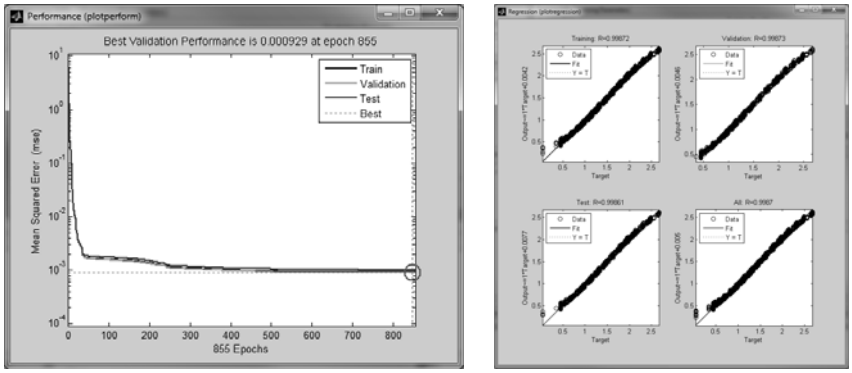


Obr. 3 Performance a regression jednej neurónovej siete pre všetky štyri kĺby

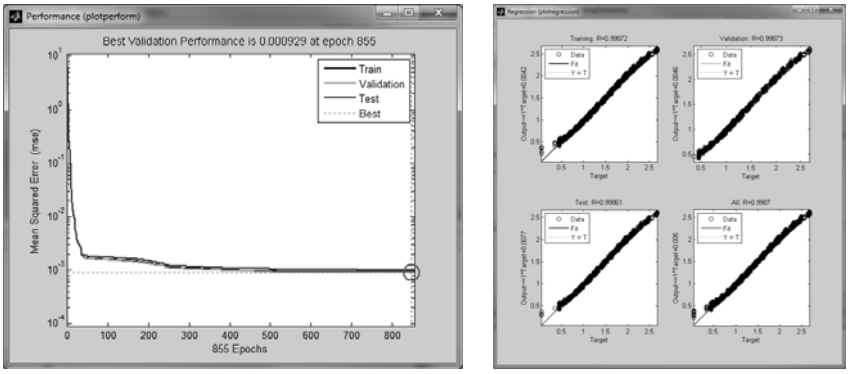
Nasledovalo učenie štyroch samostatných neurónových sietí - pre každý kĺb jedna neurónová sieť. Na vstupe boli zadefinované tri neuróny (vzorky - súradnice x,y,z) a na výstupe iba jeden neurón (jeden kĺb – stupeň voľnosti). Výsledky učenia sú zobrazené na Obr.4 – Obr.7.



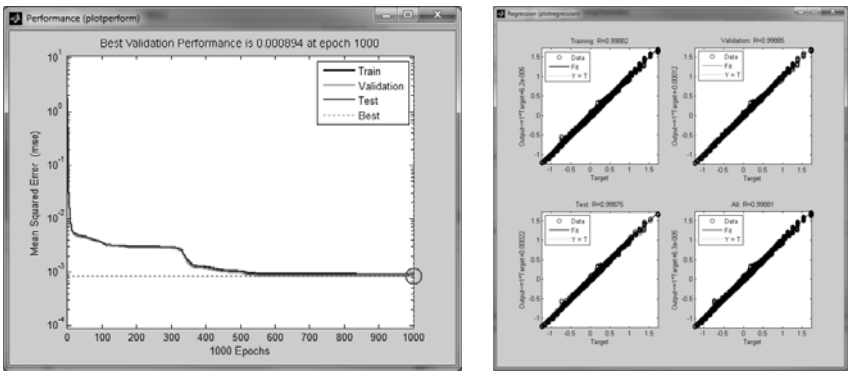
Obr. 4 Performance a regression neurónovej siete pre prvý kĺb



Obr. 5 Performance a regression neurónovej siete pre druhý kĺb

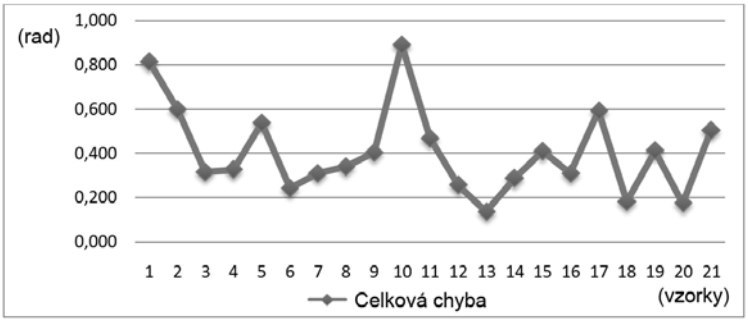


Obr.6 Performance a regression neurónovej siete pre tretí kĺb



Obr.7 Performance a regression neurónovej siete pre štvrtý kĺb

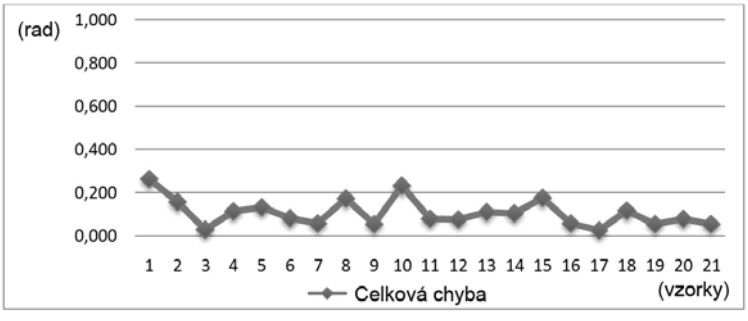
Podľa vyššie uvedených výsledkov môžeme pozorovať rozdiel v presnosti učenia sa neurónovej siete pri nižšom počte výstupov (parametre performance a regression). Pre presné overenie správnej funkcie a výslednej presnosti neurónovej siete sme zvolili vybranú časť učiacej množiny vzoriek, pri ktorých sme počítali celkovú chybu. Keďže počítame inverznú kinematickú úlohu pre robotické rameno so sériovo zapojenými stupňami voľnosti, chyby na každom stupni voľnosti sa sčítavajú do celkovej chyby. Na Obr.8. je uvedená celková odchýlka neurónovej siete pri prepočte všetkých stupňov voľnosti.



Obr.8 celková chyba jednej neurónovej siete pri prepočte všetkých stupňov voľnosti (maximálna chyba – 0,889 rad, priemerná chyba – 0,405 rad, minimálna chyba – 0,137 rad)

Na Obr.9 sú uvedené výsledky celkovej chyby pre štyri samostatné neurónové siete, pričom každá má za úlohu prepočet iba jedného stupňa voľnosti. Pozn.: aj v tomto prípade sa ako celková chyba rovná súčtu všetkých chýb každého stupňa voľnosti.

V obidvoch grafoch je použitá rovnaká mierka z dôvodu ľahšieho pozorovania rozdielov celkovej chyby.



Obr.10 - celková chyba štyroch neurónových sietí na prepočet štyroch stupňov voľnosti (maximálna chyba – 0,263 rad, priemerná chyba – 0,105 rad, minimálna chyba – 0,025 rad)

Z výsledkov je zrejmé, že pri použití jedinej neurónovej siete a štyroch samostatných neurónových sieťach je rozdiel v presnosti určenia polohy jednotlivých kĺbov trojnásobný, dokonca pri niektorých vzorkách až štvornásobný, pričom čas prepočtu je približne rovnaký.

III. ZÁVER

V príspevku bola riešený problém určenia polohy jednotlivých kĺbov sériového robotického ramena pri inverznej kinematickej úlohe riešenej pomocou umelej neurónovej siete.

Pre riešenie boli použité dva prístupy. V prvom prístupe bola navrhnutá neurónová sieť pre určenie štyroch parametrov polohy jednotlivých kĺbov sériového robotického ramena.

V druhom prístupe boli navrhnuté 4 analogické neurónové siete pre určenie jediného parametra každého kĺbu.

Z porovnania dosiahnutých výsledkov vyplynulo, že pokiaľ chyby v prvom prístupe dosahovali celkovú chybu približne 0,4 rad, tak v druhom prístupe bola celková chyba približne 0,1 rad. Dosiahnuté výsledky potvrdzujú vynikajúce aproximačné vlastnosti dopredných neurónových sietí a výhodnosť rozdelenia zložitého problému na paralelné riešenie viacerých jednoduchších úloh, pričom bola dosiahnutá podstatne vyššia presnosť a čas riešenia sa podstatne nezmenil.

POĎAKOVANIE

Vďaka za podporu projektu VaV operačného programu, Centrum excelentnosti výkonových elektronických systémov a materiálov pre ich komponenty II. No.OPVaV-2009/2.1/05-SORO, ITMS 26220120046 financovaného Európskym fondom regionálneho rozvoja (ERDF).

LITERATÚRA

[1] L. Wei, H. Wang, and Y. Li , "A new solution for inverse kinematics of manipulator based on neural network," Machine Learning and Cybernetics, Vol. 2, pp. 1201–1203, 2003.
[2] Guez, A., Ahmad, Z., 1989. A Solution to the Inverse Kinematic in Robotics Using Neural Network Processing. IJCNN89. vol. II, 299- 304.
[3] B. Choi and C. Lawrence, "Inverse kinematics problem in robotics using neural networks," National Aeronautics and Space Administration, Lewis Research Center, Cleveland, 1992.
[4] Krose, BJA. and van der Smagt, PP., 1993. An Introduction to Neural Networks. 5th ed. Chap. 7 Robot Control, University of Amsterdam.
[5] D. Pham, M. Castellani, and A. Fahmy "Accountability learning the inverse kinematics of a robot manipulator using the Bees Algorithm," 6th IEEE International Conference on Industrial Informatics, pp. 493–498, 2008.
[6] E. Gallaf , "Multi-fingered robot hand optimal task force distribution: Neural inverse kinematics approach," Robotics and Autonomous Systems, Vol. 54, No. 1, pp. 34–51, 2006.
[7] R. Köker, C. Öz, T. Çakar, and H. Ekiz , "A study of neural network based inverse kinematics solution for a three-joint robot," Robotics and Autonomous Systems, Vol. 49, pp. 227–234, 2004.
[8] J. Guo and V. Cherkassky, "A solution to the inverse kinematic problem in robotics using neural network processing", International Joint Conference on Neural Networks, Vol. 2, pp. 299–304, 1989.
[9] Lee, S., Kil, RM., 1990. Robot Kinematic Control based on Bidirectional mapping Neural Network. IJCNN90. vol. III, 327-335.
[10] Lou, YF., Brunn, P., 1999. A Hybrid Artificial Neural Network Inverse Kinematic Solution for Accurate Robot Path Control. Proceedings of the I MECH E Part I Journal of Systems & Control in Engineering, 213: 23-32.
[11] Guez, A., Ahmad, Z., 1989. Accelerated Convergence in the Inverse Kinematics via Multilayer Feedforward Networks. IJCNN89. vol. II, 341-344.
[12] Martin, P., Millan, JDR., 2000. Robot arm reaching through neural inversions and reinforcement learning. Robotics and Autonomous Systems, 31: 227-246.
[13] Timko, Jaroslav – Žilková, Jaroslava – Girovský, Peter: Modelovanie a riadenie elektrických pohonov s využitím neurónových sietí. Košice, 2009

Concept of the Virtual Laboratory for the Assembling Technologies in Electronics

Dominik DEMETER, Juraj BANSKÝ

Department of Technologies in electronics, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Slovak Republic

dominik.demeter@tuke.sk, juraj.bansky@tuke.sk

***Abstract* — This contribution is concerning with the concept of a virtual laboratory oriented to education of technological disciplines, especially in the field of technologies in electronics. It is a brief overview of the problems, which can be solved by a virtual laboratory and also contains the overview of possible ways, how to include the real technological equipment into the virtual laboratory through remote access.**

***Keywords* — e-Learning, remote access, virtual laboratory**

I. INTRODUCTION

New information technologies have affected every field of the human life. A striking impact of the internet and computers can be seen in the field of education. New, attractive education materials end tools are created, which can effectively supplement the classical printed manuscripts and help students to understand the lessons. In the field of technological disciplines, the most used e-Learning education materials are flash animations and digital video records, which help students to understand the technological principles.

II. MOTIVATION FOR CREATING A VIRTUAL LABORATORY

Creation of a virtual laboratory is usually motivated by different reasons, but the most of them are following one goal: to open up the laboratories for more users as the real laboratories and to support the education process. A well designed virtual laboratory can solve the space and capacity problems of the real technological laboratories of the education institute; help with experiments made with explosive or dangerous materials; easily include to the education process some experiments, which are not available in the real laboratories (e.g. because of missing equipment, expensive materials, etc.); improve the efficiency of the laboratory equipment usage during some time-consuming experiments and processes; guarantee the clean environment in the technological laboratories; keep the cost of the laboratory experiments on a reasonable level.

III. DIFFERENT TYPES OF VIRTUAL LABORATORIES

Two basic types of virtual laboratories can be recognized. On the one side, there are virtual laboratories, which are created independently on any real technological laboratories, on the second side, a virtual laboratory can acts as a user interface for real laboratory equipment in laboratories and provide remote access to them.

A. Virtual laboratories based on simulations

This type of virtual laboratory exists only in the virtual space and its features and possibilities are limited only by the knowledge of the implementers. All features must be programmed manually and must model the laboratory equipment and its functions. The reality and accuracy of the experiment depends on the quality of implementation of the physical laws and principles. The quality of the measurement data and the work with the laboratory is in continuous proportion with the quality of implementation.

A big advantage of this solution of a virtual laboratory is a fact, that it can be used at the

same time practically by unlimited number of persons. The only limiting factor of the availability is the performance of the application server, which can be optimized for the expected number of users.

B. Virtual laboratories connected with real laboratory equipment

This type of virtual laboratory provides a user interface for remote access of the real laboratory equipment for making remote laboratory experiments on real equipment in the real life environment. The measured data have the same accuracy as the data get from the classical laboratory experiment because of the same conditions. The only difference is the fact that the user do not need to be physically present in the laboratories. This guarantees the cleanness of the laboratories.

This solution has also some negatives. It is a challenge to find the best solution, how to substitute the operator in the laboratory and how to automatize every steps of the experiments (e.g. replacing the materials, manipulation with the final sample, moving the sample from one equipment to other, etc.). These tasks can be done by robots, but it is very expensive. This is the main reason, why only some experiments can be accessed remotely and the most of virtual laboratories of this type are oriented mainly to some prepared measurements. Other disadvantage of this type of virtual laboratories is the fact, that the equipment can be accessed and controlled only by one person, which limits the overall number of users, who can use the virtual laboratory. An exclusive access to the selected equipment can be easily granted in form of a reservation system, but it is limited to only a small number of users, which is less, than in case of simulated virtual laboratories. Other issue, which must be resolved, is to reset the equipment and the experiment to the default state and give access to a clean experiment for every user.

IV. Overview of methods, how to connect laboratory equipment to virtual laboratory

Based on the communication interfaces, we can recognize three types of laboratory equipment. Most of the modern laboratory equipment has the option of Ethernet interface through which it can be connected to the computer network. Unfortunately, this is a trend only in the last few years and the communication interface of the older equipment is limited only to the serial communication, which is sometimes missing, too. At the Department of technologies in electronics all of these types of equipments will be connected to the virtual laboratory.

A. Laboratory equipment with Ethernet interface

Laboratory devices equipped with Ethernet communication interface is easy to connect to the virtual laboratory and use remotely. In most cases, the manufacturer optimizes the equipment for remote access, including the software tool, which can easily communicate with the equipment through the internet. Most modern equipment has a built-in web server, through which they can be controlled. The manufacturer also take available a programmers guide, how to integrate their real technological equipment to a complex software solution.

B. Laboratory equipment with serial interface

Most of the older equipment has the ability to communicate through the serial interface. This is used mostly for measurement data digitalization or for controlling the devices through the computer, but with some modifications, it can be also used for remote access.

The laboratory devices equipped with serial interface can be connected to the computer network by serial to Ethernet converters (e.g. XPort from Lantronix), which converts the serial communication to Ethernet communication. The manufacturer of the converter provides a software solution (e.g. Com Port Redirector from Lantronix), too, which must be installed on the user's computer, and take the remote equipment available like it was connected directly to the client's computer. A big disadvantage of this solution is the fact, that the license conditions of the most of the software tools prohibit installing the same software on more computers and de facto, it is impossible to use the hardware by more users. Sometimes, the software tools are protected by a special hardware key witch's presence is controlled during the work.

Today, when the internet connection is reliable and fast enough, it is possible to remotely access the computers connected to the equipment and through them use the laboratory equipment. The disadvantage of this solution is that the computer and the equipment must be turned on all the time, what generates additional costs. The advantage is that the user can effectively use the software solutions delivered with the hardware, including the hardware key. It is also possible to give remote access to the users (mainly students and teachers) for some expensive, specialized software tools, like CAD and CAM tools and graphical editors. This

helps to increase the efficiency of their usage, because the remote access is available not only at the lessons, but from home, too. In this way, the equipment and software tools can be used 24 hours a day, 7 days in the week.

C. Laboratory equipment without any communication interface

A small number of older equipment does not have nor serial interface, neither Ethernet interface. It is very hard to connect these devices to the virtual laboratory, but it is not impossible. Equipment of this type requires individual solution. Mostly we need to digitalize the analog user interface by some microcomputer, which will control the device and which can be controlled remotely through the internet. Sometimes, the cost of this solution is very high and is not reasonable.

V. DIGITAL VIDEO AS INTEGRATED PART OF THE VIRTUAL LABORATORY

For remote using of equipment or controlling the experiments it is important to see, how things are working. This can be easily made by on-line digital video streaming from the technological laboratories. On-line video streaming is also useful for education purposes to show the students detailed views about the steps of the technological experiments.

Digital video records organized in a video archive can acts also as an education material. Our virtual laboratory will be education-oriented and will contain digital video records about the steps of the laboratory experiments. The video records will be created mostly in our laboratories and supplemented by audio track, which will explain the technological processes. We have also available some flash animations, which can be included to the digital video and which can help the students to understand the principles of the technological processes.

VI. CONCLUSION

At the Department of technologies in electronics we are working on creation of a virtual laboratory for assembling technologies in electronics, which will support our students in their study and will provide remote access to the laboratory equipment and computers with specialized software. Our students will be able to use the expensive CAD tools for preparing their term and diploma work. Naturally, the virtual laboratory will be available for the employee of the department for research work and controlling of the technological experiments.

ACKNOWLEDGMENT

This paper was developed with support of the project "Centrum excelentnosti integrovaného výskumu a využitia progresívnych materiálov a technológií v oblasti automobilovej elektroniky" (Centre of Excellence of the Integrated Research & Exploitation the Advanced Materials and Technologies in the Automotive Electronics"), ITMS 26220120055, that is co-financed from Structural Funds EU ERDF within Operational Program Research and Development OPVaV-2009/2.1/03-SORO and preferred axis 2 Support of Research and Development.

REFERENCES

[1] XPort, Embedded Ethernet Device Server, http://www.lantronix.com/device-networking/embedded-deviceservers/xport.html
[2] Com Port Redirector, http://www.lantronix.com/device-networking/utilities-tools/comport-redirector.html

Design and Comparison of Different Methods of Measurement of Low Pass Filter for Ultra Wide-Band Application

[1]*Alena PIETRIKOVÁ,* [1]*Kornel RUMAN,* [1]*Igor VEHEC,* [2]*Pavol GALAJDA*

[1] Department of Technologies in Electronics, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Slovak Republic
[2] Department of Electronics and Multimedia Communications, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Slovak Republic

[1]alena.pietrikova@tuke.sk, [1]kornel.ruman@tuke.sk, [1]i.vehec@tuke.sk, [2]pavol.galajda@tuke.sk

***Abstract* — This paper deals with design, simulation and comparison of different methods of measurement of a low pass (LP) filter for Ultra Wide-Band (UWB) radar applications. Paper focuses on various methods of measurement overall response and technological possibilities of production that are used in UWB filters. This paper demonstrates the design of LP filter realized by the *Filter Solution 2011* from Nuhertz as well as its simulation realized by *Ansoft Designer* from ANSYS. Simulated and different methods of measurement results of insertion loss and return loss response of LP filter for UWB radar are presented. This paper assesses the suitability of hydrocarbon ceramic laminate substrate for the production of LP filter for UWB frequency area. It examines potential manufacturing possibility of such filter based on Low Temperature Ceramics (LTCC) technology too. There is antialiasing LP filter presented, made from hydrocarbon ceramic laminate and FR4 substrate for particular laboratory UWB radar.**

***Keywords* — Low pass filter, UWB, insertion loss, return loss, hydrocarbon ceramic laminate substrate, FR4**

I. INTRODUCTION

UWB radar systems are used in a variety of applications (autoelectronics, medical application, detection of subjects behind the wall, ..). UWB applications are well-known for usefulness in localization or in communication services as well as in sensor implementations. However, big challenge is design, simulation and construction of a quality filter for removing unwanted frequencies. In particular, design, simulation and construction of LP filters for UWB radar applications with stop band frequency 2.5 GHz, is the challenging problem [1].

Nowadays, there are a variety of Computer aided Design (CAD) programs for simulation and design of filters for high frequency (HF) area in the market. When the filters are designed for HF area, there's necessary of S parameters (Scattering parameters) simulation. The LP filter was designed by software *Filter Solution 2011* from Nuhertz, the worldwide leader in low and HF filter synthesis and analysis. The design was exported to *Sonet Designer* from ANSYS and simulated with using planar EM (Electromagnetic) analysis [2, 3].

In the UWB area, the dielectric properties of substrate have a major impact on the quality and stability of the filter. Each substrate at the market has various parameters. This is the reason why every substrate is not proper for UWB area and choosing the appropriate substrates should pay close attention.

Those with higher dielectric constants are more suitable for lower frequency applications. The FR4 (Flam Resistant based on epoxy and glass – FR4) laminate is the most common electronic carrier for circuits operating below the microwave frequency. Beyond 1 GHz, the laminate properties (like ε_r, tan δ, roughness, etc…) of the FR4 become nonlinear. It becomes difficult to precisely characterize the FR4 at the desired microwave frequency. Beyond 3 GHz, the use of the FR4 is not recommended anymore, in the case of unacceptable attenuation. This is the reason why we chose HF circuit material - substrate based on hydrocarbon ceramic laminate (ROGERS RO4003C) [4]. Non-insignificant parameter for UWB application is the value of dissipation factor Tan (δ). Very important role in this area has LTCC too.

The main goal is to apply advanced digital signal processing and RF filter design methods for solving acceptable attenuation in defined band.

II. FILTER DESIGN

A. Substrate

For correct design and simulation it is necessary to know the parameters of the substrate. We choose hydrocarbon ceramic laminate as major substrate (ROGERS RO4003C) and FR4 only for demonstrates affect of substrate on transmission characteristic. Specification of used substrate based on hydrocarbon ceramic laminate (ROGERS RO4003C) is in Table 1.

Table 1.
Rogers RO4003C substrate specification

Property	Symbol	ROGERS RO4003C
Dielectric Constant	ε_r	3,55
Dissipation Factor Tan	δ	0,0027
Dielectric Height	h	0,020" (0,508 mm)
Resistivity Compared to Copper	ρ	1
Metal Thickness	M	½ oz. (17µm)
Volume Resistivity	-	1.7 x 1010 (MΩ•cm)
Surface Resistivity	-	4.2 x 109 (MΩ•cm)
Moisture Conductivity	-	0.71 (%)
Melting temperature Tg	-	>280 (°C DSC)

Low Dissipation Factor Tan (δ) allows hydrocarbon ceramic laminate material to be used in many applications where at higher operating frequency are the conventional laminates (FR4) circuit boards limited.

The temperature coefficient of dielectric constant is among the lowest of any circuit board material, and the dielectric constant is stable over a broad frequency range (Figure 1). This makes it an ideal substrate for broadband applications [5].

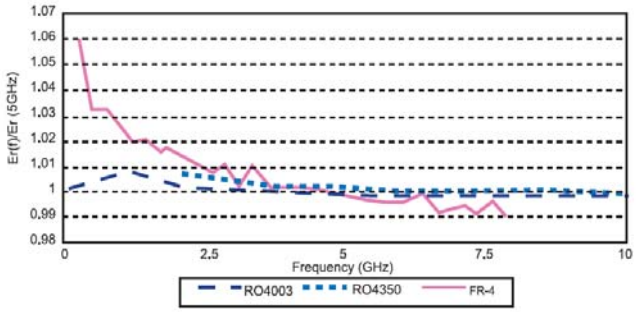


Fig. 1 Change in dielectric constant due to the increasing frequency [5].

From the Figure 1 it can be concluded that the best stability of dielectric constant in the whole frequency range from the comprised substrates has hydrocarbon ceramic laminate RO4003C. On the other hand we can see that FR4 rapidly change its dielectric constant mainly in the range of frequency from 0 to 3 GHz. Both ROGERS material prove stabile properties in the whole frequency range.

B. Initial Design

A lot of passive filters used in UWB are known as Gaussian, Bessel, Butterworth, Chebyshev I, Chebyshev II, Elliptic etc. [2]. After simulation we decided to choose the filter type Elliptic, which requires attenuation and impulse response. Based on the results of the simulations, the optimum minimum order of 6 was determined, in which filters meet the required min. attenuation of -40 dB at frequencies from 2.5 GHz to 15 GHz. The insertion loss magnitudes of Elliptic filter oscillate up and down within a user-specified boundary. The higher order of the filter, the more oscillations occur. These oscillations are known as Pass Band Ripple [2]. The pass band ripple of 0.05 dB is used. Layout of the LP filter, designed with the help of *Filter solution 2011* is shown in the Figure 2a.

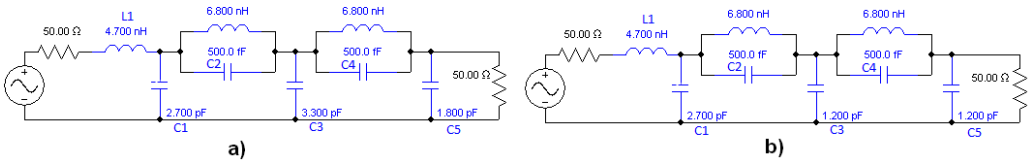


Fig. 2 The scheme of the LP filter a) original scheme designed by Filter solution 2011 b) scheme with modified capacitors C3 and C5

The LP filter input and output are matched to 50 ohm characteristic impedance. The 50 ohms choice is a compromise between power handling capability and signal loss per unit length, for air dielectric [6].

C. Simulation

One of the major problems which are solved in the area of LP filter for UWB radar applications (frequencies from 2.5 GHz to 15 GHz) is necessity for simulation of S parameters. The quality of the filter is affected by many factors such as substrate material, size of the components, distance among components, dimensions of microstrip lines, bends etc. Solving these problems served for the correct construction of a LP filter. The design from *Filter Solution 2011* was exported to *Sonet Designer* from ANSYS. The final design of LP filter for ROGERS RO4003C substrate is shown in Figure 3, including the ports, microstrip lines, bends, SMD passive components (size 0201) and distance between components. Simulated performance data and filter layout are derived from this data.

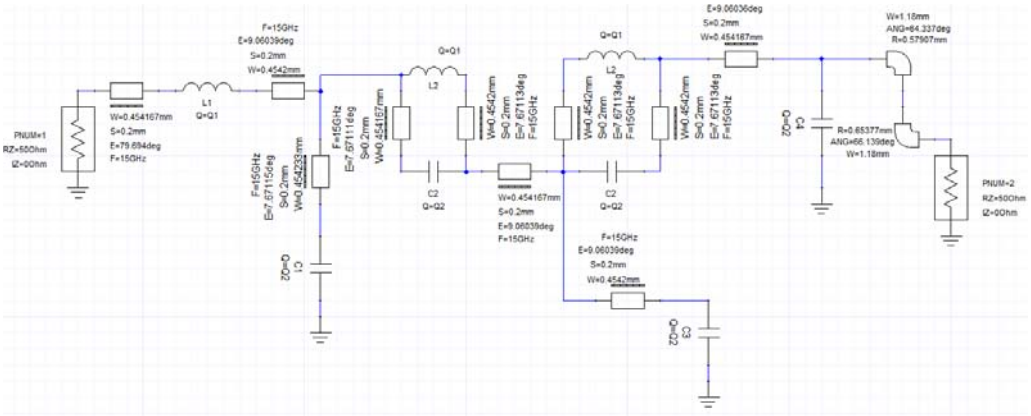


Fig. 3 The final design of LP filter for ROGERS RO4003C designed with the help of Sonet Designer from ANSYS.

This layout data were used for simulation with planar EM (Electromagnetic) analysis. The insertion loss and return loss of LP filter for ROGER RO4003C substrate after EM simulation and ideal insertion loss are shown in the Figure 4.

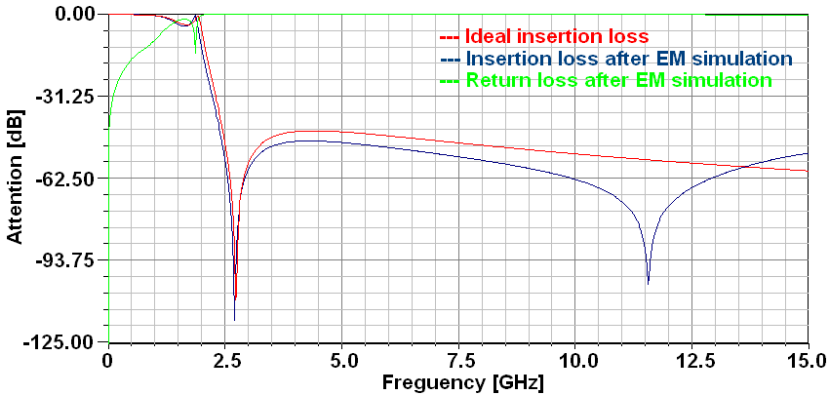


Fig. 4 Simulated results of insertion loss and return loss of LP filter for ROGER RO4003C substrate

Figure 4 shows magnitude of change in the S21 transmission characteristics between ideal insertion loss and insertion loss after EM simulation. EM simulation was purposely made up to

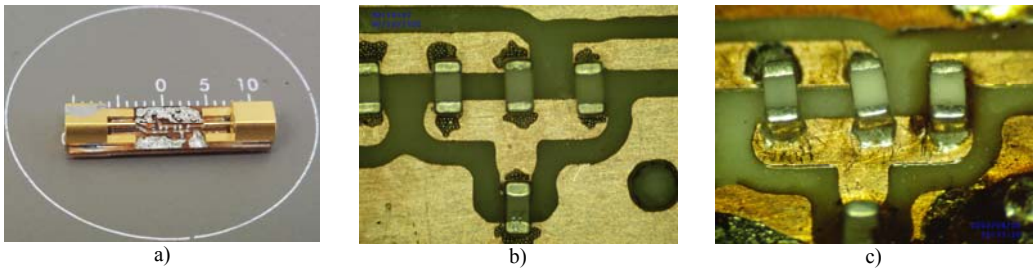
15 GHz to verify the correct function of the filter at high frequencies. After simulations it was necessary to change the value of C3 and C5 components to provide minimum attenuation -40dB at frequencies from 2.5 GHz to 15GHz. The modified scheme is shown in the Figure 2b.

EM analysis result indicates that the insertion loss and return loss meets the design criteria.

III. RESULTS

A. *Filter construction*

Final prototype of filter was made of ROGER RO4003C and RF4 substrate using the CIF CNC drilling & milling machine (Fig. 5a). Surface-mount device (SMD) components and SubMiniature version P (SMP) connectors were soldered to the filter by Vapour Phase Soldering (VPS) process immediately after leadfree SAC solder paste deposition and components placing. We used VPS device and temperature profile, presented in [7].The SMD of 0201 size were used for miniaturization as well as for active suppression of parasitic effects in high-frequency range. The Figure 5b shows samples just before soldering and Figure 5c after VPS.



a) b) c)

Fig. 5 LP filter a) Final prototype of LP filter b) samples before VPS c) samples after VPS

B. *First technique of measurement*

The first technique of measurement of LP filters was performed on the spectrum analyzer (Fig. 6). As the input signal, of LP filter we used signal from generator of pseudorandom M-sequences MX-9 developed on TU Ilmenau. MX-9 contains 9 shift registers (9^2 -1 samples = 511 samples) and has been clocked to 6 GHz.

The Figure 7 shows pseudorandom sequence (power spectrum of signal in dBm) on spectrum analyzer without LP filter.

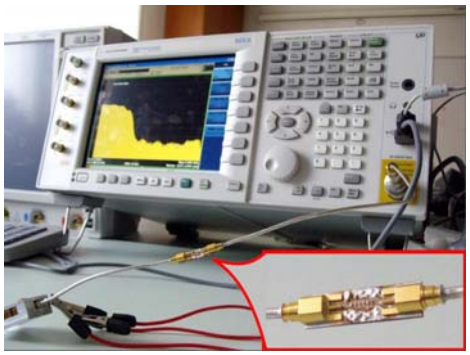


Fig. 6 The first control measurement of LP filters on the spectrum analyzer

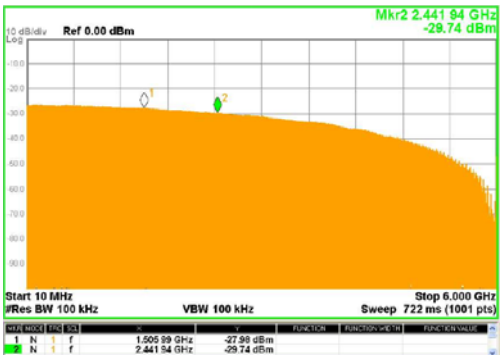


Fig. 7 Power spectrum of pseudorandom sequence on spectrum analyzer

The Figure 8 shows the power spectrum of pseudorandom sequence after filtering signal with LP filter made of ROGER RO4003C substrate. The insertion loss of the LP filter, measured by spectrum analyzer, match with simulated results for frequencies from 0 GHz to 6 GHz.

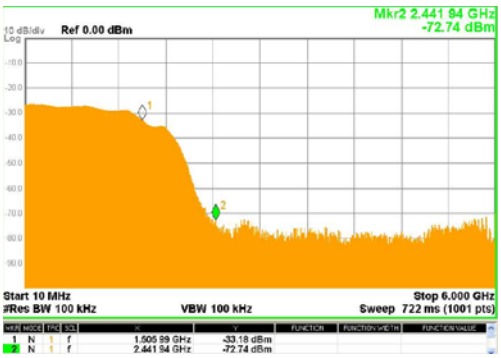


Fig. 8 Power spectrum of pseudorandom sequence after filtering signal with LP filter made of RO4003C substrate



Fig. 9 Power spectrum of pseudorandom sequence after filtering signal with LP filter made of FR4 substrate

The Figure 9 shows the power spectrum of pseudorandom sequence after filtering signal with LP filter made of FR4 substrate. This LP filter was not optimized for this FR4 substrate (we expected worse result than in ROGER RO4003C substrate) but this measurement demonstrates impact of substrate (its dielectric height, dielectric constant, etc.) on transmission characteristic.

It is obviously that for evaluate quality of filter this method of measurement is not sufficient and it can be used only as control measurement during development when the vector network analyzer is not available.

C. Second technique of measurement

The second technique of measurement of LP filters was performed on the Rohde & Schwarz vector network analyzer. Using this vector network analyzer the insertion loss and return losses from both samples were measured. Output signal from vector network analyzer were stored into PC for signal processing using *Matlab*. Comparison of simulated and measured insertion loss for LP filter made of ROGERS RO4003C and FR4 are shown in the Figure 10.

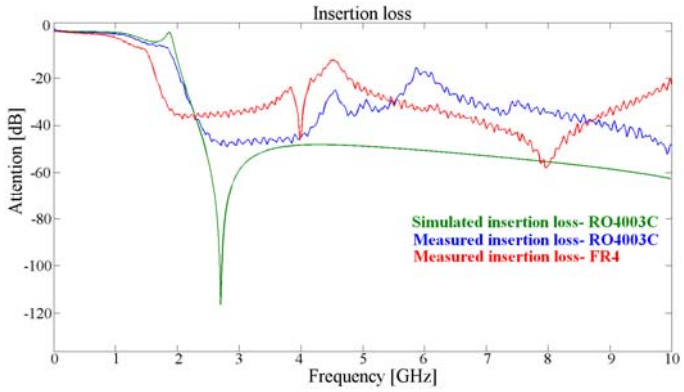


Fig. 10 Comparison of simulated and measured insertion loss of LP filter made of Rogers RO4003C and FR4 substrate

Figure 10 shows that the insertion loss of the LP filter made of Rogers RO4003C doesn't match with its simulated results. Insertion loss of LP filter made of Rogers RO4003C meet the initial conditions only for frequencies from 0 GHz to 4,5 GHz but the transmission characteristic of LP filter is still acceptable for use in the UWB applications. It is possible to use this filter as antialiasing LP filter in UWB devices such as IQ demodulator presented in [8] or part of system for Through Wall Moving Target Tracking by M-sequence UWB Radar presented in [9].

Figure 10 also shows that the insertion loss of the LP filter made of FR4 doesn't match with simulated results. The reason was that simulation of LP filter was tailored and optimized for ROGERS RO4003C substrate (not FR4). This measurement demonstrates the effect of physical and material characteristic of substrate (dielectric height, dielectric constant, etc.) on transmission characteristic. Even the fact that LP filter made of FR4 substrate doesn't match with the simulated cure is interesting because from 5.5 GHz it has better course in compare with ROGERS RO4003C. But because of the stop band frequency, which has moved from 2.5 GHz to 2 GHz, the LP filter made of FR4 substrate doesn't meet the design criteria. On the other hand we can say that FR4 substrate is more proper (stabile) for frequency range from 5.5

GHz to 8.5 GHz.

Comparison of simulated and measured return loss for LP filter made of ROGERS RO4003C and FR4 are shown in the Figure 11.

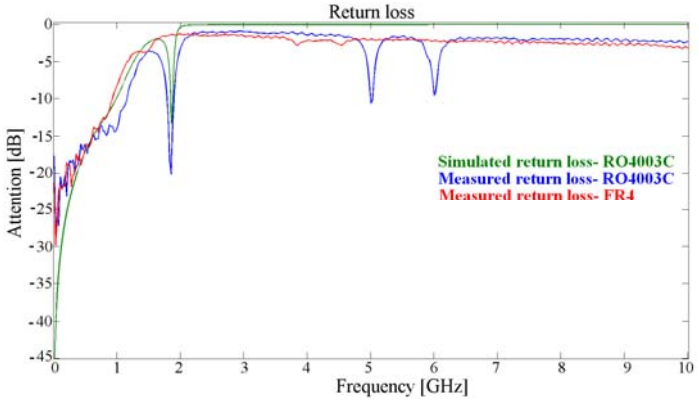


Fig. 11 Comparison of simulated and measured return los of LP filter made of Rogers RO4003C and FR4 substrate

Return loss of LP filter made of ROGERS RO4003C and FR4 substrate is sufficiently consistent with simulated results for both samples. Acceptable profile of insertion loss results in worse return loss but it was counted with this situation in simulation.

IV. Conclusion

The LP filter with stop band frequency 2.5 GHz based on hydrocarbon ceramic laminate (ROGERS RO4003C) and FR4 substrate was designed, simulated and constructed. Comparison between simulated and measured results of insertion loss and return loss were presented. We also demonstrate impact of substrate quality (its dielectric height, dielectric constant, etc.) on transmission characteristic and suitability of hydrocarbon ceramic laminate substrate for UWB area. It has been shown that for evaluated quality of filter measurement with spectrum analyzer isn't sufficient and it is necessary more complex measurement on vector network analyzer. Transmission characteristics of LP filter made of ROGERS RO4003C are acceptable for use in the UWB applications. In the future work, we will improve vias for connection with ground and shielding around LP filter. This project was basis for manufacturing such filter based on LTCC multi-layer technology.

Acknowledgment

This paper (50%) was developed with support of the project "Centrum excelentnosti integrovaného výskumu a využitia progresívnych materiálov a technológií v oblasti automobilovej elektroniky", ITMS 26220120055, that is co-financed from Structural Funds EU ERDF within Operational programme Research and Development OPVaV-2009/2.1/03-SORO and preferred axis 2 Support of Research and Development. This paper has been supported by foundation of national projects (50%) No. KEGA 030_TUKE_ 4/2011, VEGA 1/0298/09 as well as APVV Sk- Ro – 0025-10.

References

[1] I. J. Immoreev, and J. D. Taylor, "Future of radars," Ultra Wideband Systems and Technologies, 2002. pp. 197-199.
[2] http://www.nuhertz.com/profile.html.
[3] http://www.ansoft.com/products/hf/ansoft_designer.
[4] Nikholas G. Toledo, "Practical Techniques for Designing Microstrip Tapped Hairpin Resonator Filters on FR4 Laminates", in Communications Engineering Division, Advanced Science and Technology Institute, Diliman Quezon City, Philippines 1101.
[5] http://microboone-docdb.fnal.gov/cgi-bin/RetrieveFile?docid=1896;filename=RO4000%20%20Laminates%20-%20Data%20sheet%20.pdf;version=1.
[6] http://microwaves101.com./encyclopedia/why50ohms.cfm.
[7] Livovský Ľ., Pietriková A., Ďurišin J., "Monitoring of the temperature profile of vapour phase reflow soldering", ISSE, Budapest, 2008, ISBN 9789630649155, p. 687-689.
[8] Žiga M., Liptaj M., "Design of the I-Q Demodulator for UWB Applications", SCYR 2012 - 12th Scientific Conference of Young Researchers - FEI TU of Košice, 2012, Košice, will be published.
[9] Rovnakova J., Kocur D., "UWB Radar Signal Processing for Through Wall Tracking of Multiple Moving Targets", 7th European Radar Conference, 2010, 204-207 p.

Simulation of Stencil Printing Process

Alena PIETRIKOVÁ, Michal KRAVČÍK

Department of Technologies in Electronics, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Slovak Republic

alena.pietrikova @tuke.sk, michal.kravcik@tuke.sk

***Abstract*—this paper presents results of investigation of rheology behaviour of three solder pastes with various densities in the context of simulation of stencil printing process in electronics. Comparison of three viscosity structural models was used for description of rheological flow curve of these solder pastes samples. Solder pastes are metal alloys suspended in flux medium and spotlight of this paper is on modelling ofviscositystructural models of solder paste. Solder paste rheology is the most important parameter which affects the paste printing process, the reflow soldering performance, and wherefore the resulting solders joint quality and reliability. The rheology behaviour can be correlated to the performance of the materials during the stencil printing process. Solder pastes exhibit non-Newtonian rheological characteristics such as shear thinning i.e. the reduction of viscosity with increased shear. Viscosity structural models help to understand the rate of structural change of solder paste. The viscosity values from structural models are compared with empiric measured values and printing quality tests.These viscosity structural models can be use for solder paste stencil printing simulations, which can prevent incorrect stencil printing setting with given solder paste rheology.**

***Keywords*—Solder paste, viscosity structural model, rheology, solder paste printing simulation**

I. INTRODUCTION

The printing of solder pastes through very small stencil apertures required for flip-chip pitch sizes is expected to result in increased stencil clogging and incomplete transfer of paste to the printed circuit board pads. Nowadays it is one from the most critical point in the fluently continuous miniaturization in electronics. The most of assembly defects eventuates from improperly understanding of the effect of printing process parameters on printing performance. Solder paste is pseudo-plastic non-Newtonian fluid, which is shear thinning with rising shear rate. It is also known that solder paste exhibit a thixotropic behaviour, which is recognized by the decrease in apparent viscosity with time when subjected to a constant shear rate.

To understand rheological phenomena associated with the flow of solder pastes, it is necessary to understand time-dependent rheological behaviour of solder paste because it is crucial for establishing the relationships between the pastes structure and flow behaviour as well as for correlating the physical parameters with paste printing performance. Motivation for this study is determine parameters for viscosity structural models with dependency on temperature. Determination of the most accurate model with suitable parameters for the solder paste in a given environment will allow us to create highly accurate simulation and estimation behaviour solder paste during the printing process. Determination of the mathematical description of the behaviour of the solder paste, depending on the shear rate, is the first and most important step for the computational simulation. Result of viscosity structural model helps us to understand a wide spectrum of rheological behaviours during the printing process and influences the establishment of proper metal content as well.

II. EXPERIMENTS

Our experiments were oriented to identification of influence of various densities (% of metal content) as well as temperature and determine parameters for three different viscosity structural models. Results from empiric measurements and results from structural models were compared on printing performance of solder pastes with the aim to identify the technological condition in which solder paste has very good stencil print quality.

A. Sample Materials

One commercial grade solder paste (Paste P1 88.5%) was modified by increasing of metal

contend to 90% of metal weight (Paste P2) and 91.5% of metal weight (Paste P3). All composition were based on 96.5Sn-3Ag-0.5Cu, flux type M14, melting temperature 217 – 220°C, diameter of particles were 25 – 45 μm. SnAgCu powder particles characterised by the identical physical properties in compare with commercial paste was admixed.

B. Viscosity Structural Models

Many rheological modelshave been established to describe the viscosity. Only some are intended to describe non-Newtonian fluids. Three simply viscosity models can be used to estimate the shear thinning property of non-Newtonian materials, Cross Model, Sisko Model and Power Law Model[1], [3]. They provide better estimates at different sections of a viscosity versus shear rate diagram.

- **Cross Model** is one of the most widely used models for describing non-Newtonian fluids. Searched parameters are K_1 and *n1*.

$$\frac{\eta-\eta_\infty}{\eta_0-\eta_\infty}=\frac{1}{1+K_1\gamma^{n1}} \qquad (1)$$

- **Sisko Model** for a description of non-Newtonian fluids at high shear rate. Searched parameters are *K2* and *n2*.

$$\eta=\eta_\infty+K_2\gamma^{n2} \qquad (2)$$

- **Power Law Model** is the simplest of models to describe the behaviour of non-Newtonian fluids. Searched parameters are $K_3$and *n3*.

$$\eta=K_3\gamma^{n3-1} \qquad (3)$$

Over the empiric measured data of viscosity was estimated all unknown parameters. The obtained parameters were set to the formulas in models and the resulting values were compared with measured viscosity values. For determination of unknown parameters its necessary simplify mathematic formulas of structural methods to linear configuration. Since this are a relatively simple mathematical formulas, linearization of these formulas is quite simple. After model formulas linearization can find unknown parameters by solving a linear regression. Since the linear regression solution strongly depends on the range of input parameters, these calculations were carried out in multiple variations. The aim of these calculations was to find unknown parameters to be the most accurate description of the flow curve samples. After finding the appropriate parameters for each structural model, these models were compared with each other, along with empirical data.

III. Computed Parameters and Their Accuracy

For these three types of paste samples were computed parameters for viscosity structural models in Tab 1. As written above, the values of search parameters were obtained by solving linear regression. Since the resulting parameter values are highly dependent on the choice and quantity of input data, several computations were performed to obtain these parameters.

Table 1
Computed unknown parameters of both samples for all three viscosity structural models

Sample	Parameters	Cross Model	Sisko Model	Power Law Model
P1 at 24°C	**K**	11.621	1216.66	487.5232
	n	1.61	-1.6047	0.338
P2 at 28°C	**K**	1.708	9455.49	625.4292
	n	2.34	-2.3377	0.3184
P3 at 30°C	**K**	1.645	9945.49	702.5841
	n	3.25	-2.4245	0.3002

Comparisons of the deviations of the values obtained with these three models with the actual measured values for all samples of solder pastes were public at our previous studies[2][3]. In short we can say that all three models have high accuracy in compare with empiric measured

values. And Power Law Model has the higher accuracy and also its largest deviation is less than 3%. In our study the rheological properties of solder pastes, the accuracy of the results of the three viscosity structural models are very similar for other temperatures and metal content studied three solder pastes. But here there is no space for their presentation.

IV. STENCIL PRINTING PROCESS SIMULATION

The purpose of the stencil printing process is to deposit solder paste at precise locations on the metal pads of printed circuit boards (PCBs) in preparation for the positioning of surface mount components and reflow soldering. This process involves pushing the solder paste over the surface of a stencil ahead of a squeegee blade. The paste is then forced into apertures in the stencil, assisted by the high pressure generated in the region close to the blade (Fig. 1). The apertures coincide exactly with the pattern of metal pads on the PCB. The stencil is separated from the PCB following the pass of the squeegee blade. Since the adherence of the solder paste to the metal pads is stronger than that to the aperture walls, free-standing paste is left on the PCB in contact with the metal pads in the required locations. Following the printing process, surface mount components are placed on the PCB and reflow soldering occurs, in which heat is applied so that the solder particles melt and subsequently re-solidify to form a bond.

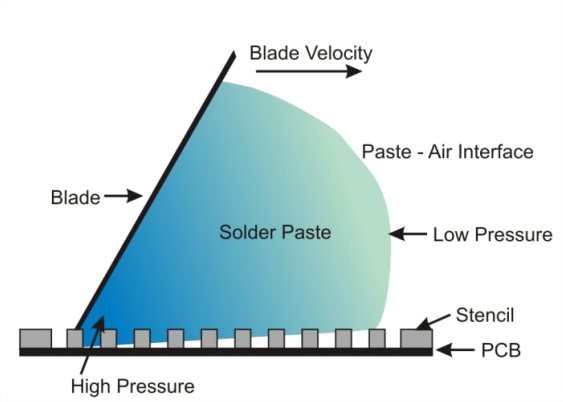


Fig.1Scheme of stencil printing process

Various analysis techniques, including analytical models and numerical simulations, can be use to study different aspects of the printing process in an attempt to improve the understanding of the critical parameters that affect the printing quality. Such studies have made an important contribution to the understanding of solder paste behaviour but work is still needed for further development of these methods to provide more reliable guidelines for process optimization. Owing to the extremely complex interactions involved, determining the specific effect of varying any one of the governing process parameters alone is extremely difficult without taking into account all other critical factors. The understanding of the printing process has therefore developed from piecing together information from various different analysis methods.

The boundary conditions of micro-structural studies of the printing process require realistic data based on the macroscopic flow conditions in the whole mass of solder paste, particularly along the stencil surface. Macroscopic flow characteristics such as pressure, velocity, strain rate and viscosity are extremely difficult, if not impossible, to measure experimentally during the squeegee stroke but can be predicted with the use of analytical or computational fluid dynamics methods. In addition to providing boundary conditions for small-scale studies, such data may also help to understand experimentally observed correlations between process operating parameters, such as blade angle and velocity, with the resulting print quality.

The most often used methods for simulation of stencil printing process in electronicsare Riemer's Analytical Method that is based on mathematical analysis as well as Computational Fluid Dynamics Method (CFD) that is based on graphical-computational analysis[5].

A. Riemer'sAnalytical Method

Riemer' s analytical model [6], developed originally for the analysis of off-contact ink screen printing, provides relationships that give the pressure distributions in ink ahead of a moving squeegee blade. This model has been applied to the solder printing process, providing a

valuable guide to the flow conditions encountered. The pressure distribution along the stencil surface is given by

$$P = \frac{1}{r}\left(\frac{2sin^2\theta}{\theta^2 - sin^2\theta}\right)\eta v \qquad (4)$$

where$P$ is the hydraulic pressure, η is the apparent viscosity, v is the squeegee speed, θ is the squeegee angle and r is the distance from the blade tip. The analytical model is applicable in 2D space simulation for limit number of points investigated. Solder paste is non-Newtonian fluid showing thixotropic properties and therefore the relationship between pressure and shear rate is dependent on the applied stress time. Relationship between pressure and shear rate, which is necessary to determine viscosity at investigated point is too complex to describe in detail and will be published in our next articles.

B. Computational Fluid Dynamics Method (CFD)

Since the pressure along the stencil surface predicted by Riemer's model is proportional to the reciprocal of the distance from the blade tip, the theoretical value tends to infinity as the blade is approached. This cannot be the case in reality. Beyond some point approaching the blade tip, the analytical model becomes increasingly invalid as, in reality; the pressure is dissipated by physical effects, whether by deformation of solid surfaces, the forcing of paste under the blade or stencil, three dimensional effects or any other physical reason. This raises the issue of the proximity to the blade tip in which Riemer's analytical model can realistically represent the behaviour of a fluid while ignoring such physical effects.

The movement of the paste is modelled using the PHYSICA multiphysics computational modelling framework [5], [7]. This software utilizes the finite volume method for flow analysis and provides an environment suited to modifications such as the inclusion of user defined parameters and material behaviour.The motion of the paste is analysed in this study as a two-dimensional, steady state, single-phase problem. The paste is treated as a continuum for which the rheological and physical properties are averaged over a large area.

Owing to the discretization of space representing the region of paste in the CFD analysis (Fig.2), the pressures obtained at the elements closest to the blade tip are a finite value. The CFD results can only follow the analytical curve up to a certain proximity to the blade tip, depending on the density of the grid in this region. Inspection of the results indicates that the density of the grid used in this study produces a curve very close to that of the analytical model at distances greater than 1.0×10^{-4} m from the blade tip, which is the same as the possible width of a fine-pitch stencil aperture.Closer results could be obtained nearer the blade tip by increasing the density of the grid in this region. However, treatment of a solder paste as a continuum is not suitable for the analysis of the physical behaviour of a solder paste at this scale since the size of the solder particles, ranging from approximately 1.0×10^{-4} m to 2.0×10^{-5} m in diameter, are larger than the elements of a sufficiently fine grid. A microstructural analysis is therefore required at this scale, which is capable of accurately representing the motion of individual solder particles and their interaction with each other as well as the surrounding solid component surfaces.

Pressure contours close to the blade tip are shown in Fig. 3 for non-Newtonian CFD studies.The pressure contours in the non-Newtonian case follow a smooth arc in the region near to the blade tip, whereas further away the contours become refracted.Lines in the figure are the points of equal pressure and their value increases linearly. Pressure the highest increase is on the tip of a blade. Refracted lines on the opposite side represent rollingof the solder paste.

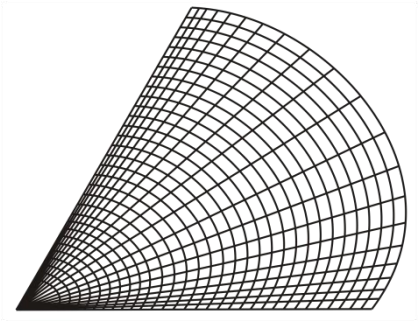


Fig.2 Simply grid model for CFD

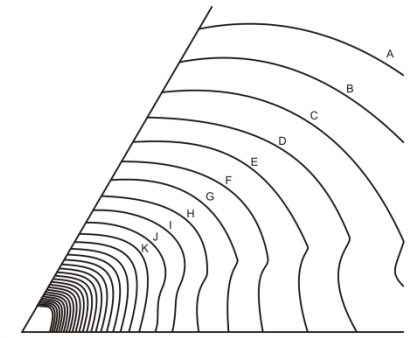


Fig.3 Pressure contours for non-Newtonian Paste properties

V. CONCLUSION

More future effortsare required between the analyses at the macro and micro scales to result in realistic models that provide useful data to solder paste manufacturers and the electronics packaging industry. Non-Newtonian models and simulations, analytical models can provided the best indication of flow behaviour and pressure distributions in a roll of solder paste.

ACKNOWLEDGMENT

This paper (50%) was developed with support of the project "Centrum excelentnostiintegrovanéhovýskumu a využitiaprogresívnychmateriálov a technológií v oblastiautomobilovejelektroniky", ITMS 26220120055, that is co-financed from Structural Funds EU ERDF within Operational programme Research and Development OPVaV-2009/2.1/03-SORO and preferred axis 2 Support of Research and Development. This paper has been supported by foundation of national projects (50%) No. KEGA 030_TUKE_ 4/2011, VEGA 1/0298/09 as well as APVV Sk- Ro – 0025-10.

REFERENCES

[1] TA Instruments RHEOLOGY APPLICATIONS NOTE, "Rheology Software Models Flow," http://www.tainst.com, 2008.
[2] A. Pietrikova, M. Kravcik, "Investigation of Rheology Behavior of Solder Paste," The 35th International Spring Seminar on Electronics Technology ISSE 2012, Bad Aussee 2012.
[3] M. Kravcik, "Preparation of Viscosity Structural Models for Solder Pastes," Poster 2012, Praha 2012.
[4] M. Kravcik, A. Pietrikova, "Printing Performance of Solder Paste Dependent on Temperature and Density," 12th Scientific Conference of Young Researchers, Kosice 2012.
[5] T. A. Nguty, N. N. Ereke, "Modeling the effects of temperature on the rheology of solder pastes and ¯ux system", JOURNAL OF MATERIALS SCIENCE: MATERIALS IN ELECTRONICS 11, 2000.
[6] Riemer, D. Analytical model of the screen printing process: Part 1. Solid State Technol., August 1988,
[7] PHYSICA Multi-Physics Computational Modelling Software, available at http://physica.ac.uk.

Influence of Ultrasonic Wedge Bonding Parameters on Resistivity

Igor VEHEC, Alena PIETRIKOVÁ, Kornel RUMAN

Department of Technologies in Electronics, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Slovak Republic

i.vehec@tuke.sk, alena.pietrikova@tuke.sk, kornel.ruman@tuke.sk

***Abstract* — the paper deals with investigation of influence of ultrasonic wedge bonding parameters on the resistivity of AlSi-1% bonds bonded on the thick-film pads with different metallurgy system (Ag/Pt and Ag/Pd). The resistivity of wire bonds were measured by 4-point (Kelvin) method. The realized experiments show that wire bonds resistivity depends on the setting of bonding parameters (force, power, time) and the minimal resistivity is in the range of optimal bonding parameters.**

***Keywords* — ultrasonic AlSi-1% wire bonding, wedge bonding, resistivity of wire bonds, thick-film technology**

I. INTRODUCTION

AlSi wire is classical material that is attracting for more years in wire bonding technology due to its advantages in comparison with gold or copper wire. This paper presents an achievement of ultrasonic wedge bonding with 25 μm AlSi-1% wire on Ag/Pt and Ag/Pd plated alumina substrate at ambient temperature. The making of quality interconnection is prerequisite for sufficient reliability of electronic component, particularly in critical and demanding conditions. Impact of bonding parameters (force, power and time) to the electrical resistivity of the wire bond is evaluated in this paper.

Wire bonds can be divided into few categories according to various criteria. There are known thermo-compress (T/C), ultrasonic (U/S) and thermo-sonic (T/S) bonding divided according to physical principles on which they are bonded. Approximately 93% of all semiconductor packages are manufactured using ball bonding method, while wedge bonding is used to produce about 5% of all assembled packages. Another way of classification is according to the shape of bond and used bonding tool/tip on ball and wedge bonding. Wedge bonding is named based on the shape of its bonding tool. In this technique, the wire is fed at an angle usually 30-60^{o} from the horizontal bonding surface through a hole in the back of a bonding wedge. Very important is wire material which should be on the basis of Al, Au or Cu. Aluminium is often alloyed with 1% Si to provide a strengthening mechanism. At room temperature, 1% silicon exceeds the solubility of silicon in aluminium by a factor of 50, which leads to silicon precipitation. The number and the size of the silicon precipitates are dependent on the cooling rate from higher temperatures. Slower cooling rates result in more precipitation and large nonuniform silicon nodules, while faster cooling rates do not allow sufficient time for silicon precipitation resulting is uniformly dispersed nodules. Silicon grain size can affect wire ductility; the second phase can become a potential nucleation site for fatigue cracks. There is necessary not only set of correct bonding parameters but also choosing of the correct bonding tip according to the used wire material. Bonding tips may have different shapes that affect the quality of bond and bond length. Optimization of bonding process requires include assessment of a number of factors that can affect on the bond quality (die, packaging, bonding wires, wire bond tools, bonding parameters, etc.). The typical method for assess of bond quality is based on destructive (destructive pull/shear test, etc.) and non-destructive (nondestructive pull test, visual inspection, etc.) methods.

II. RESISTIVITY OF THE WIRE BONDS

A. Principle of wire bond resistivity measurement

Bonded materials include wire and pad metallization. To obtain a good bond, the electrical and mechanical characteristics of the bonded materials must be taken into account wire selection includes: wire material, wire diameter, wire aspect ratio from the allowable ranges isolated. Critical wire material properties include electrical conductivity, shear strength, tensile strength (yield and ultimate), elastic modulus, Poisson's ratio, hardness, and coefficient of thermal expansion. Pad material must be selected to match the wire used. The pad material properties include: conductivity, bond-ability, activity to form intermetallic phases and Kirkendall voids, hardness, corrosion resistance, coefficient of thermal expansion. After bonding, the wirebond can be evaluated with visual method and mechanical testing, depending on the requirements and the situation that may arise from previous experiences with a particular package or mechanical techniques. Visual method uses optical microscope, scanning electron microscope (SEM), and other analytical instruments to find the undesired bonds. Wire bond resistivity measurements are very important too.

The proposed principle of measuring of wire bond electrical resistivity is shown in Fig. 1.

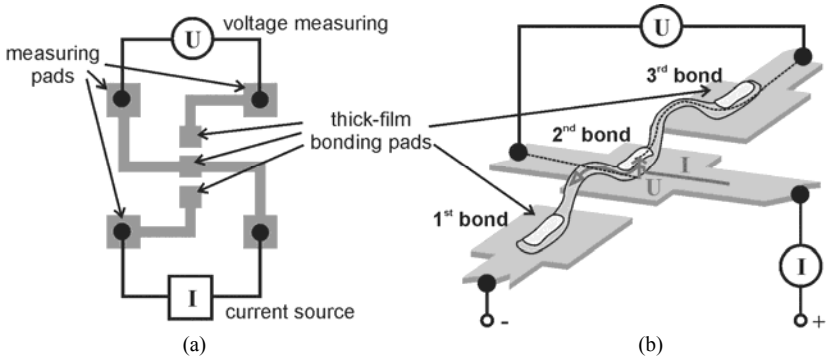


Fig. 1 The topology of thick-film conductive layer for one thick-film element for measuring of 2nd bond resistivity (a) and principle of 2nd bond measuring (b).

The "stich" interconnection consists of three bonds bonded on the thick-film pads. The resistivity of 2nd bond was measured by standard 4-point (Kelvin) method where part of the interconnection between 1st and 2nd bond served as one of the wire from current source, and part of interconnection between 2nd and 3rd bond served as one of wire for voltage measuring. The whole current supply loop consist of (thick-film conductive layer) → measured 2nd bond ← (interconnection between 2nd and 1st pad and thick-film conductive layer) where brackets distinguished the measuring wire. Analogous, the whole voltage loop consist of (thick-film conductive layer) → measured 2nd bond ← (interconnection between 2nd and 3rd pad and thick-film conductive layer). Therefore, the voltage was measured only on the close part of the 2nd bond, through which current flows from the source and where is voltage measured. Ideally it would be measured only contact resistance of 2nd bond at the interface between bonded wire and thick-film pad.

B. Samples preparation and resistivity measurement

The thick-film topology of the final conductive layer was adapted for alumina substrate with standard dimension of 2 × 2'' (50.8 × 50.8 mm) and it is shown on Fig. 2a.

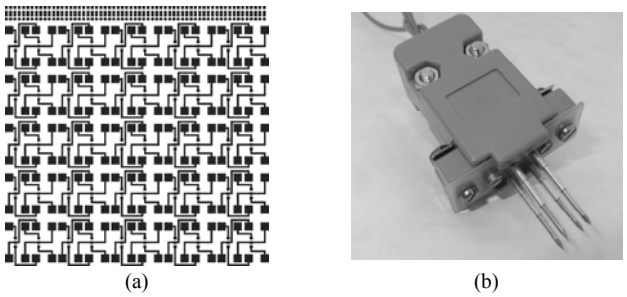


Fig. 2 The topology of final thick-film conductive layer (a) adjusted for substrate with dimension of 2 × 2'' (50.8 × 50.8 mm) and for 4-point probe with tungsten tips (b).

The basic thick-film elements are tailored for 4-point probe (Fig. 2b) and for better utilization of space on the substrate. Upper part of topology (upper line on Fig. 2a) was used only for bonding and testing of bonds by destructive pull test.

The conductive layer was screen-printed on standard alumina substrate (96% Al_2O_3) with two thick-film pastes based on different system metallurgy: 5210 TT based on Ag/Pd system and C1076 SD based on Ag/Pt system. Typical properties of these pastes are shown in Tab.1.

Table 1
Typical properties of used thick-film pastes

	TT 5210	C1076 SD
Paste system metallurgy	Ag/Pd	Ag/Pt
Viscosity [Pa.s] @ 25°C	170 ÷ 230	30 ÷ 50
Fired resistivity [mΩ/□]	30	4
Resolution (lines/spaces) [μm]	200/200	200/200
Substrate	Alumina (Al_2O_3)	
Firing profile	850°C / 60 min	

The "Kulicke & Soffa 4129 Deep Acces Wedge Bonder" was used for wire bonding. Normally, the first bond is made to the die and the second is made to the substrate. The reason is that it can be far less susceptible to edge shorts between the wire and die. By descending the wedge onto the IC bond pad, the wire is pinned against the pad surface and an U/S or T/S bond is performed. Next, the wedge rises and executes a motion to create a desired loop shape. At the Second bond location, the wedge descends, making a second bond. During the loop formation, the movement of the axis of the bonding wedge feed hole must be aligned with the center line of the first bond, so that the wire can be fed freely through the hole in the wedge. The used wedge tool was a Maxiguide type with bond length of 89 μm and the used wire with 25 μm in diameter made by Heraeus was based on soft alloy of AlSi-1%. The breaking load of this wire is 10 cN. AlSi wire was bonded in an ultrasonic bonding process at room temperature. A considerable advantage of the wedge bonding is that it can be designed and manufactured to very small dimensions, down to 50 mm pitch. The bonding parameters of wire bonder equipment were changed during bonding and subsequently influence of bonding parameters on the 2nd bond was measured. The K&S 4129 wire bonder allows adjust three typical bonding parameters in these ranges:

- ultrasonic power up to 1.3 W (resp. up to 2.5 W),
- force applied on the bond in the range 10 to 160 g,
- and duration of ultrasound action (time) in the range 10 to 200 ms.

The optimal bonding parameters for sufficient bond quality were found at the beginning of the experiment. The wire bond quality was evaluated by visual inspection as well as by destructive pull test carry out by equipment constructed on our Department of Technologies in Electronics. The determined optimal wire bonding parameters for wire bonding with AlSi-1% wire on both Ag/Pt as well as Ag/Pd thick-film pads are shown in Tab. 2. These parameters are valid for the bonding on the presented type of thick-film pastes and they weren't considered in the bonding on the die pads.

Table 2
Optimal wire bonding parameters for AlSi-1% wires bonded on thick-film pastes

Wire bonding parameter	Value
Force	~35 g
Power	~0,42 W
Time	~67 ms

The influence of wire bonding parameters on 2nd bond resistivity was measured in the next experiments. The each one selected parameter (successively: force, power and time) was changed in the small range around optimal parameter while the remaining two parameters were kept at optimum levels (Tab.2). The resistivity of bonds was measured by Agilent 34420 Micro-Ohm Meter with the measuring probe with four tungsten tip (Fig. 2b).

III. EXPERIMENTAL RESULTS

The measured dependencies of electrical resistivity of 2nd bond on force, power and time are shown in Fig.3 to Fig. 5. For each measured point there are marked the percentage of good bonds for which was not occur bond destruction during wire bonding. The measured points are evaluated as average resistivity of 20 correctly bonded bond. The individual bonds were bonded until the correct 20 bonds were bonded. Subsequently percentage of good bonds was evaluated as ratio of good bonds (20) to total number of attempt of bonds creation.

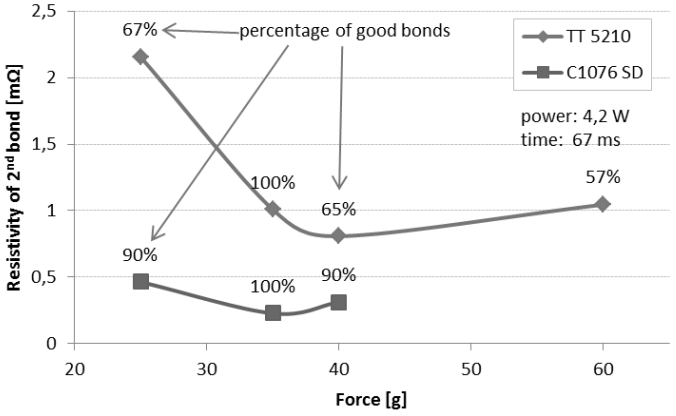


Fig. 3 Resistivity of 2nd bond as a function of applied force applied on bond.

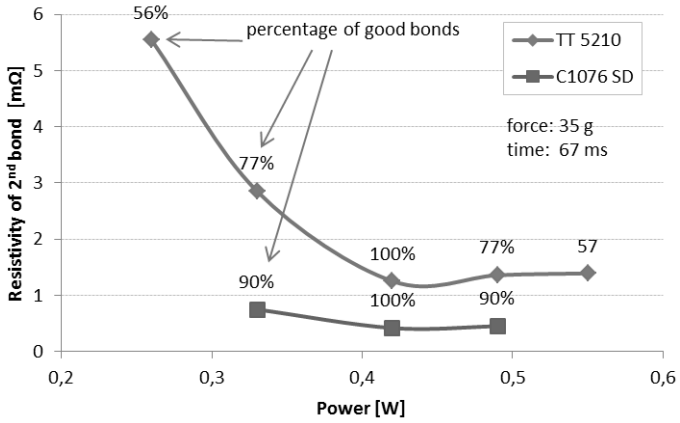


Fig. 4 Resistivity of 2nd bond as a function of ultrasonic power.

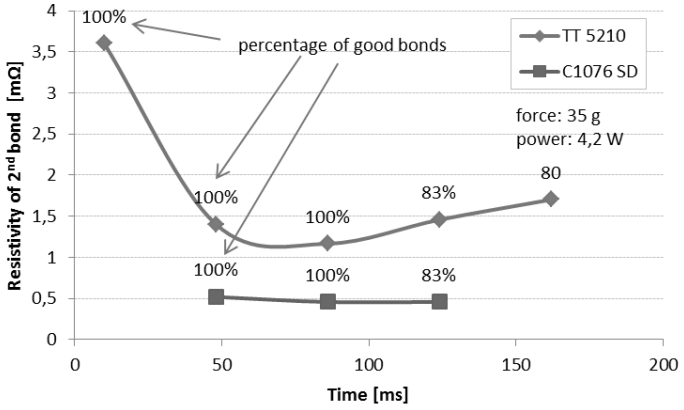


Fig. 5 Resistivity of 2nd bond as a function of duration of ultrasonic energy.

The all graphs has minimum of bond resistivity in the region of optimal parameters for both

used thick-film pastes. The increasing of resistivity out of optimal parameters can be explained either by insufficient bond creation at lover parameter levels, or by excessive deformation of bonds at higher parameter levels.

As we can see, the resistivity of bonds bonded on the pads based on Ag/Pt paste (SD C1076) is lower as resistivity of bonds bonded on the pads based on Ag/Pd paste (TT 5210) for all dependencies. It can be caused not only by own resistivity of 2nd bond, but also by lower sheet resistance of Ag/Pt thick-film paste compared to Ag/Pd paste. The resistivity could be affected by different thick-film conductive layer, too. The same screen was used for screen-printing of both pastes, but the pastes have different viscosity (Tab. 1) which could affect printability and final layer thickness and subsequently bond-ability and resistivity.

Ag/Pt paste offer wider variability of bonding parameter settings in the terms of the percentage of good joints, because the percentage at same parameters was always higher for Ag/Pt paste than Ag/Pd paste. The adjusting of time (Fig. 5) has smaller influence on percentage of good bonds as other two parameters (force – Fig. 3 and power – Fig. 4) for both types of pastes.

IV. Conclusion

The experiments show that wire bonds resistivity depends on the setting of bonding parameters (force, power, time) and on type of used thick-film paste. The minimal resistivity is in the range of optimal bonding parameters. The resistivity of bonds bonded on Ag/Pt pads are lower as bonds bonded on Ag/Pt as could be presume. The optimal parameters for AlSi-1% wire with 25 µm in diameter wedge bonding on both types pads (based on Ag/Pt and Ag/Pd pastes) are:

- force: ~35 g,
- power: ~0,42 W,
- time: ~67 ms.

These parameters are valid for the bonding on the presented type of thick-film pastes and they were not considered in the bonding on the die pads.

Acknowledgment

This paper (50%) was developed with support of the project "Centrum excelentnosti integrovaného výskumu a využitia progresívnych materiálov a technológií v oblasti automobilovej elektroniky", ITMS 26220120055, that is co-financed from Structural Funds EU ERDF within Operational programme Research and Development OPVaV-2009/2.1/03-SORO and preferred axis 2 Support of Research and Development. This paper has been supported by foundation of national projects (50%) No. KEGA 030_TUKE_ 4/2011, VEGA 1/0298/09 as well as APVV Sk- Ro – 0025-10.

References

[1] G. Harman, “Wire Bonding in Microelectronics”, 3rd Edition, New York: McGraw-Hill, 2010, ISBN: 978-0-07-164265-1.
[2] Bonding Wedge Catalog (Book style). Kulicke & Soffa, 2001.
[3] Z. Lai and J. Liu, Chapter A: Wire Bonding. In: The Nordic Electronics Packaging Guideline, 2000. http://extra.ivf.se/ngl/
[4] C. Wang and R. Sun, The Quality Test of Wire Bonding. In: Modern Applied Science, Vol. 3, No. 12, Canadian Center of Science and Education 2009, ISSN 1913-1852.

Intricate World of Electrical, Magnetic and Structural Properties of Rare Earths Metals Thin Films at Low Temperatures–the Case of Holmium and Thulium Thin Films

[1]*Ján DUDÁŠ,* [2]*Jozef BAGI,* [1]*Milan GUZAN,* [3]*Slavomir GABÁNI*

[1] Department of Theoretical Electrotechnics and Electrical Measurement, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Slovak Republic
[2] Húskova 31, Košice, Slovak Republic
[3] The Institute of Experimental Physics, SAS, Košice, Slovak Republic

[1]jan.dudas@tuke.sk, [1]milan.guzan@tuke.sk, [3]gabani@saske.sk

Abstract **— High precision electrical resistance (R) measurements were performed on a holmium (Ho) and thulium (Tm) bulk and their thin film samples in the temperature range from 4.2 K up to the room temperature. The X-ray diffraction of Ho films revealed diffraction peaks from the hexagonal closed packed (hcp) structure of Ho and those from the face centered cube (fcc) holmium dihydrid. Secondary ion mass spectroscopy showed a very homogeneous distribution of Ho in thin films. The crystal structure study of Tm films revealed diffraction peaks only from the Tm hcp crystal structure. The clearly seen "knee-like" resistance anomaly, observed in Ho bulk sample, is caused by magnetic phase transition with the Néel temperature (T_N) 128.9 K and the Curie temperature of this Ho bulk was 19 K. The R/R_0 *vs. T* anomalies connected with magnetic phase transitions as well as those caused by the H in the form of α-phase and dihydrid were observed in Ho thin films. The "hump-backed" resistance anomaly was observed in Tm bulk sample near T_N = 57.5 K. In contrast, unexpected resistance anomalies were observed for thulium thin films in the low temperature region below ≈ 60 K and near ≈ 170 K.**

Keywords **— electrical resistance, holmium and thulium thin films, low temperatures, Néel temperature, Curie temperature, SIMS, X – ray diffraction,**

I. INTRODUCTION

Thin films containing rare earth metals (RE) are studied because of their promising technical applications, e.g. RE/TM (Fe, Co) as valuable media for magnetic recording or high temperature superconductivity films for application in electronics. RE metal thin films as their constituents are still attractive because of the wealth of their physical properties, e. g. the variety of magnetic structures appearing at low temperatures. Size effects may produce relevant modifications in the known magnetic phases of RE elements because of nonuniform magnetic structure induced by the lower coordination near the surfaces. But the intricate world of RE thin films, connected with their reactivity as excellent getters for hydrogen, causes problems in interpretation of the studied properties (see e.g. [1]).

This paper demonstrates the intricate world of electrical, magnetic and structural properties of RE metal thin films at low temperatures in the case of holmium and thulium thin films.

*Holmium***,** like several other heavy lanthanide metals [2], orders magnetically by indirect exchange between the localized 4*f* moments (RKKY interaction). Below T_N = 130 K Ho orders antiferromagnetically in simple helical spin structure with magnetic moments in the basal plane. Below T_C = 19 K moments tilt out of the basal plane to form a conical structure.

Being paramagnetic at room temperature, *thulium* orders antiferromagnetically below the T_N = 56 K into the c-axis spin wave modulated spin structure (CAM), which gradually squares with decreasing temperature. It orders into the square wave spin structure with seven layer

sequence, four having magnetic spins up along the c-axis followed by three layers with spins down, below the Curie temperature $T_C = 32$ K .

II. EXPERIMENTAL

The investigated thin films were prepared by evaporation onto glass substrates. Holmium films were evaporated in ultrahigh vacuum at ~ 10^{-7} Pa, thulium thin films were evaporated in high vacuum of ~ 10^{-4} Pa. High precision electrical resistance measurements of bulk and thin film samples were performed by the conventional four-point dc arrangement using a Keithley digital programmable current source K 220 and digital nanovoltmeter K 181.

The temperature of the bulk and thin film samples in the helium cryostat was measured using calibrated Ge (from 4.2 K to 80 K) and Pt (from 80 K up to room temperature) thermometers. Film thickness was measured using the optical Tolansky method. The crystal structure of thin films was determined using X-ray diffraction with the Bragg - Brentano focusing geometry. The purity of Ho films was checked using SIMS.

III. RESULTS

a) *Holmium thin films*

Prior to the Ho thin film study, the electrical resistance of the reference Ho bulk sample was measured. The result of this measurement is illustrated in Fig.1 as the R/R_0 vs. T dependence (R_0 being the resistance at 4.2 K and R its value at the T). The residual resistance ratio value (ratio of the resistance at room temperature to that at 4.2 K) RRR=18.8. The "knee-like" resistance anomaly is clearly seen near ≈ ~ 130 K in Fig. 1, corresponding to the magnetic phase transition from paramagnetic state to the antiferromagnetic one with spiral structure. The numerical analysis of the resistance data yielded the Néel temperature $T_N = 128.94$ K and $T_C =$ 19 K (see the inserts in Fig.1).

The electrical resistance of holmium thin films (36, 60, 98, 116, 147, 196, 215 nm) was investigated in the temperature range from 4.2 K up to the room temperature and results of this investigation is illustrated in Fig. 2 for two of them with thickness of 116 and 215 nm. The 116 nm Ho thin film exhibits a "hump-backed" resistance anomaly near ≈ 120 K, which is connected with the magnetic phase transition to the antiferromagnetic spiral structure. The numerical analysis of experimental data yielded the value $T_N = 119.5$ K. The qualitatively other result was obtained in 215 nm thin film of holmium. The R/R_0 vs. T dependence shows anomaly near ≈ 120 K, which resembles to the "knee-like" R/R_0 vs. T anomaly observed for basal plane of holmium single crystal. Moreover, a resistance minimum was observed at 10 K (indicated by arrow T_{min} in Fig. 2) and the slope´s change of the R vs. T dependence at ~ 170 K (indicated with T_a).

The results of the crystal structure study are illustrated in Fig. 3 for 116 nm Ho film. The diffractogram contains the diffraction peaks at 33.65 ^{0}, 37.2 ^{0} and 38.65 ^{0} caused by the hcp structure of Ho and the peak at 34.9^0 is caused by the fcc structure of holmium dihydrid HoH_2.

The diffraction pattern of 215 nm thin Ho film, illustrated in Fig. 4, is quite different, it exhibits two peaks caused by the hcp structure of holmium and a large peak at 35^0, which is caused by the HoH_2.

The RE metal thin films readily react with hydrogen to form the dihydrides with the fcc structure [1], [3]. The positions of expected HoH_2 diffraction peaks was evaluated by means of the Bragg law using X-ray radiation of wavelength $\lambda = 0.17902$ nm and the lattice parameter of the fcc structure of HoH_2 a =0.5140 nm [4]. The diffraction peak caused by (111) plane of the fcc should appear at $2\Theta = 35^0$ position (Θ is the Bragg angle). The position of the observed fcc diffraction peak in 215 nm thin Ho film is in coincidence with the evaluated value, thus we assume, the holmium dihydrid is present in this film.

The depth profiles SIMS spectra taken from the 196 nm Ho film are displayed in Fig.5. These results show the film has homogeneous surface. The depth profile signals from HoH_2, Ho_2O_3, Ho_3O_4 were present together with signals from other impurities. Nevertheless, comparing the R vs. T curves of our Ho bulk sample with those of 116 nm we see that the amount of HoH_2 and other contaminants was very low and they did not influence qualitatively the observed R vs. T dependence of this film.

b) *Thulium thin films*

The electrical resistance of thulium bulk sample, investigated as a reference in the temperature range from 4.2 K up to the room temperature, showed the „hump-backed" resistance anomaly observed near ≈ 60 K, which is typical for the magnetic phase transition from paramagnetic state to the antiferromagnetic one with CAM spin structure. The residual resistance ratio value of Tm bulk sample was RRR=142. Numerical analysis of the

experimental data yielded a sharp deep minimum in the *R* vs. *T* curve giving the Néel temperature T_N = 57.5 K.

The resistance of thulium films in the thickness range from 25 nm to 115 nm was also investigated in the temperature interval from 4.2 K up to the room temperature and Fig. 6 is illustration for 115 nm thin film. In contrast to the Tm bulk results, we observed *R* vs. *T* anomalies in two temperature regions, for three films with thicknesses 25, 31 and 115 nm below ≈ 60 K and near ≈ 170 K. The anomalies below ≈ 60 K include a plateau in *R vs. T* curve in the interval ΔT ≈ 16 K below ≈ 50 K for the 115 nm thin film, indicated by the arrow T_{a1}, a sudden increase in resistance value with increasing temperature near ≈ 20 K for 31 nm thin film and even a resistance minimum at ≈ 9 K for the 25 nm thin film. All Tm films exhibited the resistance anomaly near ~ 170 K, indicated by the arrow T_{a2} in Fig. 6.

The crystal structure study of Tm films using X-ray diffraction showed the diffraction peaks caused only by hcp Tm and these peaks are caused by the preferential orientation of the crystallites.

IV. Discussion

The unexpected resistance anomalies in thin films of Ho and Tm, observed at low temperatures, which are are quite different in shape from that of the bulk sample, are caused by other mechanisms than those in the bulk.

The rare earths absorb hydrogen easily and form solid solutions and/or hydrides [5]. Their phase diagrams consist, in general, of basic parts: metal, metallic solid solution (α-phase) with the H atoms inserted in the tetrahedral intersticies of the host-metal lattice and the equally metallic dihydride phase. Hydrogen concentration influences the magnetic, transport and structural properties.

As for the Ho 215 nm thin film, we assume, that the larger content of HoH_2 in this film is the reason for the "weaker " resistance anomaly near ≈ 120 K of this film, caused by magnetic phase transition, compared with those of 116 nm thin Ho film.

The observed resistance anomalies in Ho and Tm films are probably caused by the absorbed hydrogen, which is always present in the bell-jar [6]. The H solubility at low temperatures is in Ho much smaller than in Tm and the excess solid solution of H precipitates in the form of dihydrid [5], [7]. Dihydrid was detected by the X–ray in our Ho films, but not in the Tm films. Thus the resistance anomalies in Tm films at temperatures below ~ 60 K could be interpreted as caused by the hydrogen present in solid solution (α-phase) in these films whereas in the case of Ho films by both – solid solution of H and HoH_2 phases.

As for the resistance anomalies at 170 K, it has been found that the electrical resistivity of some rare earths (Tm, Er, Lu, Sc) with H exhibits an anomaly near ~ 170 K, which is attributed to short–range ordering of hydrogen atoms coupled into H – H pairs between the two nearest–neigbour tetrahedral sites along the c-axis of the hcp host lattice (see e.g. [8]). We assume, that the observed slope´s change of the *R* vs. *T* dependences observed in our Ho and Tm films near 170 K could be interpreted as caused by short-range H-H pair ordering.

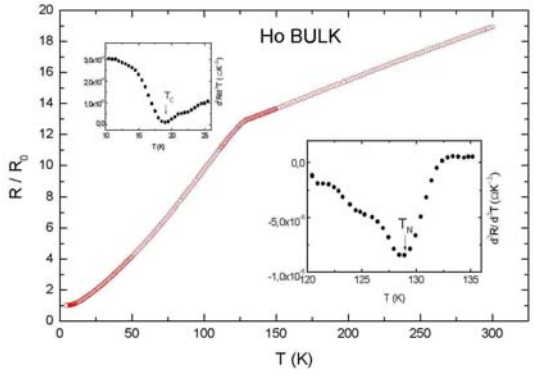


Fig.1. Temperature dependence of the resistance ratio R/R_0 of Ho bulk sample in the temperature range from 4.2 K to 300 K. The inserts show the temperature derivative of resistance vs. temperature curve near the Néel temperature T_N (right bottom) and near the Curie temperature T_C (left upper corner).

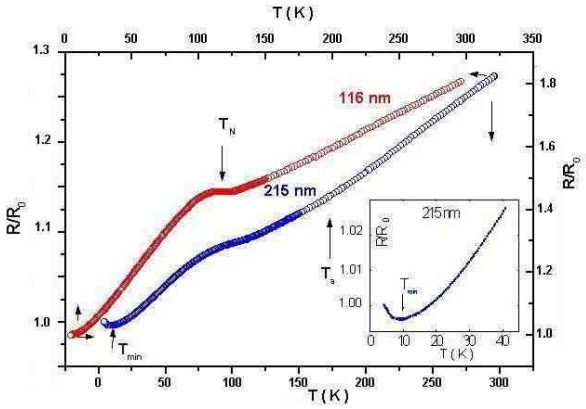


Fig.2. The resistance ratio R/R_0 vs. T curves of two Ho thin films with thickness of 116 nm and 215 nm in the temperature range from 4.2 K to 300 K. The insert shows more clearly the resistance minimum at 9 K -T_{min}. The T_a indicates the slope`s change in the R/R_0 vs. T curve at ~ 170 K of the 215 nm Ho thin film.

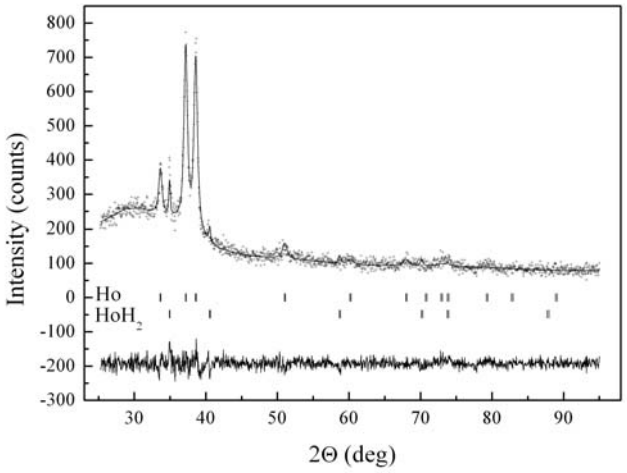


Fig.3. Rietveld refinement plot for 116 nm Ho film. The upper curve includes experimentall points overlayed by the calculated solid line. The lower curve shows the difference between the observed and calculated intensities. Positions of all possible Bragg reflections for the two phases are indicated by vertical marks in the middle

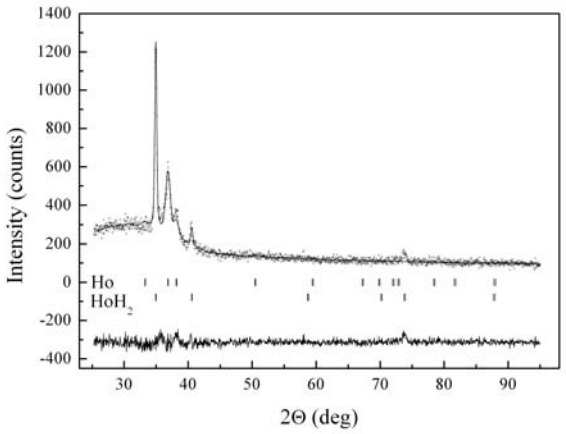


Fig.4. Rietveld refinement plot for the 215 nm thin holmium film

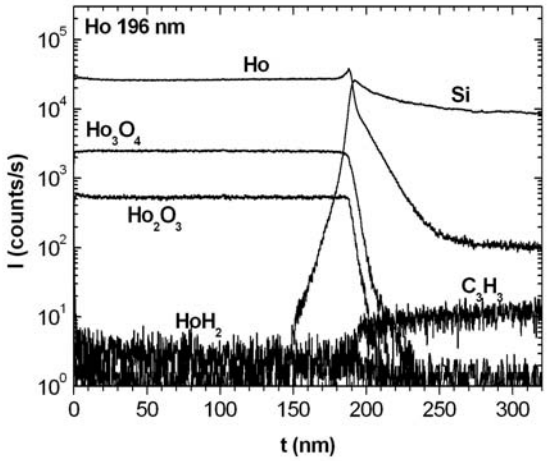


Fig. 5. The SIMS depth profile of 196 nm thin Ho film with spectral lines of Ho, Ho_3O_4, Ho_2O_3, HoH_2, C_3H_3 and Si (from the glass substrate). The lines of other impurities (see the text) are omitted because of better lucidity.

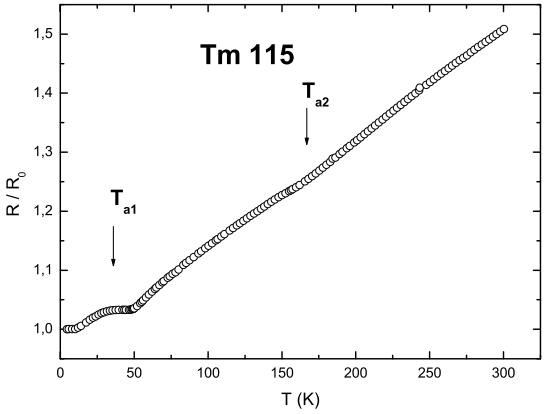


Fig.6. Resistance ratio vs. temperature of a 115 nm thulium thin film in the temperature range from 4.2 K to 300 K. Resistance anomalies below 60 K are indicated by the arrows T_{a1} and T_{a2}.

V. CONCLUSIONS

We have investigated the electrical resistance of holmium and thulium bulk and thin film samples in the temperature interval from 4.2 K up to the room temperature, crystal structure of Ho and Tm films by X-ray diffraction and purity of Ho films by SIMS spectroscopy.

It follows from this study:

1. Thin films of Ho and Tm are very sensitive to the hydrogen present in the vacuum bel-jar and in the parent material, which forms the solid solution (α phase) and excess hydrogen precipitates in the form of dihydrid.
2. The X-ray diffraction study of Ho thin films revealed also the presence of holmium dihydrid in these films.
3. The depth profiles SIMS spectra taken from the 196 nm Ho film show the film has homogeneous surface. The content of impurities was very low and they did not influence the transport properties of the films
4. The X-ray diffraction study of Tm films (HV) showed diffraction peaks caused only by the hcp structure of Tm.
5. Observed types of *R vs. T* anomalies in thulium films (HV) below ≈ 60 K are caused by the hydrogen in the form of solid solution and resistivity minimum at ~ 10 K in Ho 215 nm thin film is caused as superposition of the α phase and HoH_2.
6. Observed slope´s change of the resistance vs. temperature curves of Ho and Tm films near ~ 170 K are probably caused by the short–range ordering of hydrogen atoms coupled into H – H pairs between two nearest-neighbour tetrahedral sides along the c-axis of the hcp host lattice.

REFERENCES

[1] J.Dudáš, INTRICATE BEHAVIOUR OF RARE EARTH METAL THIN FILMS AT LOW TEMPERATURES, Mercury-Smékal Pu. House, Košice, 2001, 125 p.
[2] W. C. Koehler in *Magnetic Properties of Rare Earth Metals,* R. J. Elliott, Ed., Plenum Press, London,1972, pp.81-128.
[3] M. Gasgnier in *Handbook on the Physics and Chemistry of Rare Earths*, K. A. Gschneidner, Jr. and L. Eyring, Ed., North Holland, Amsterdamm1995, pp. 105-206.
[4] M. Gasgnier et al., Rare earth hydrides and rare earth oxides in and from thin films of rare–earth metals, J. Less – Common Metals, **38** (1974), 131.
[5] P. Vajda, Hydrogen in rare-earth metals, including RH_{2+x} phases in *Handbook on the Physics and Chemistry of Rare Earths*, K. A. Gschneidner, Jr. and L. Eyring, Ed., North Holland, Amsterdamm1995, pp. 207-291.
[6] O. S. Rajora, A. E. Curzon, On the effect of substrate temperature, deposition rate and annealing on the electrical resistivity of thin yttrium films, J. Materials Science Letters **3** (1984), 627.
[7] P Vajda, Private communication, 2004.
[8] P. Vajda, J. N. Daou, J. P. Burger, G. Hilscher and N. Pillmayr, Influence of hydrogen in solid solution on the magnetism of monocrystalline thulium, J. Phys. Cond. Matter **1** (1989), 4099

Diagnostika napájacej sústavy automobilu na báze internetu

Ján MOLNÁR, Tibor VINCE

Katedra teoretickej elektrotechniky a elektrického merania, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

jan.molnar@tuke.sk, tibor.vince@tuke.sk

***Abstrakt* — Tento článok sa zaoberá problematikou diagnostického systému napájania automobilu na báze internetu. V článku je navrhnutý algoritmus využitia Internetu ako meracej zbernice. Návrh pozostáva z novej koncepcie automatizovaného meracieho reťazca, ktorý diagnostikuje a následne vyhodnocuje napájací systém automobilu. Takto navrhnutý systém poskytuje nové doposiaľ nepreskúmané možnosti analyzovania skrytých a ťažko odhaliteľných chýb vznikajúcich pri používaní automobilu.**

***Kľúčové slova* — meranie, automobil, Internet, Expertný systém, GSM, Wifi.**

I. ÚVOD

Vzdialené meranie cez Internet sa stáva novým štandardom v meracej technike, ktorému sa začína venovať veľká pozornosť hlavne v poslednom období. Internet sa stáva bežným prostriedkom pre získavanie informácii, ale nachádza svoje uplatnenie aj v meracej technike. Jeho hlavnou výhodou je to, že umožňuje vzdialene meranie na veľké vzdialenosti. Ďalším veľkým benefitom Internetu je to, že namerane údaje môžu pochádzať z rôznych zdrojov, ich počet nie je žiadnym spôsobom limitujúci. Týmito výhodami sa Internet začína stavať súčasťou meracieho reťazca, ktorý nachádza uplatnenie v rôznych odvetviach priemyslu.

V niekoľkých posledných rokoch sa automobilový priemysel stáva veľmi významným odvetvím, a to nielen pre slovenskú ekonomiku. Spomínaný rozvoj so sebou priniesol aj nové potreby diagnostikovania a následného odhalenia rôznych skrytých nedostatkov a chýb, ktoré vznikajú pri bežnej prevádzke automobilov. Takto vznikla aj požiadavka na vývoj diagnostického systému napájania v automobile, ktorý je určený pre meranie porúch nabíjacej sústavy. Namerané dáta by sa za pomoci Internetu mali odoslať do expertného systému, ktorý tieto dáta vyhodnotí a pokúsi sa odhaliť príčinu danej poruchy.

Nakoľko sa jedná o nový koncept merania, treba si uvedomiť, že automobil nie je možné pripojiť k žiadnej metalickej komunikačnej sieti. Preto sa pri tomto meraní využívajú len bezdrôtové technológie založené na mobilnej sieti GSM a bezdrôtovej sieti Wifi. Obe siete umožňujú meracej aparatúre umiestenej v automobile, pripojiť sa k Internetu, a tým zabezpečiť real–time meranie. Je potrebné si však uvedomiť, že aj vyššie spomenuté technológie trpia určitými nedostatkami, ktoré môžu mať za následok stratu nameraných údajov. Preto je potrebný aj dlhodobý zber dát, aby sme boli schopní eliminovať určíte výpadky.

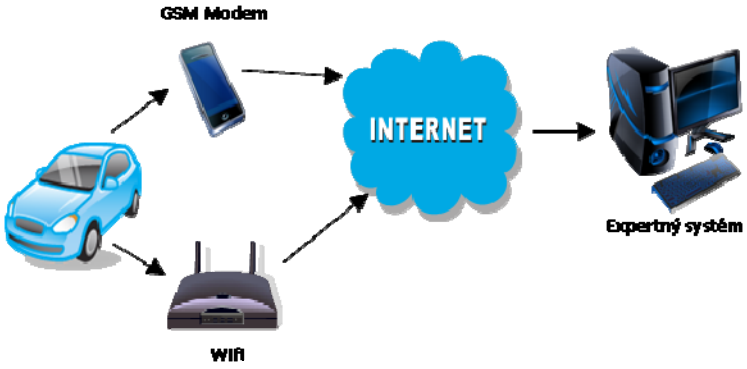
II. NÁVRH MODELU MERACIEHO REŤAZCA

Vzhľadom na rozmáhajúci sa automobilový priemysel, aj v tomto odvetví je veľké množstvo problémov, ktorých riešením môže byť práve vzdialené meranie. Mnohým poruchám a poškodeniam v automobile by bolo možné predísť prevenciou alebo včasným diagnostikovaním poruchového stavu, prípadne určením samotnej poruchy. V princípe by sa jednalo o meranie zvolených údajov v automobile, prenos údajov do centra na spracovanie a vyhodnotenie situlacie.

Dnešný prístup spočíva v tom, že v pravidelných intervaloch, v rámci predpísaných kontrol v autoservise je auto napojené na počítač, kde sa prevedú všetky merania. Bolo by však oveľa efektívnejšie, ak by bolo možné, v prípade podozrenia na poruchový stav alebo odklon od optimálnej prevádzky, preniesť údaje do počítača s expertným systémom, ktorý by daný stav vyhodnotil. Meranie by prebiehalo počas bežnej prevádzky vozidla, čím by sa uľahčilo aj

identifikovanie vzniku neperiodických chýb, ktorých identifikácia je najťažšia, keďže v servise sa nemusí podariť danú chybu vyvolať a odstrániť. Týmto spôsobom by bol možný prenos údajov vo chvíli výskytu danej chyby.

Architektúrou daného pripojenia je zobrazená na Obr. 1. Navrch sa zaoberá „zbernicou“ – prvkom meracieho zariadenia, ktorý je zodpovedný práve za prenos nameraných dát k vyhodnocovaciemu zariadeniu – expertnému systému a jeho vlastnosťami.



Obr. 1. Návrh architektúry vzdialeného merania v automobile.

Ako je zrejmé z obrázka, automobil by mal v sebe integrovaný GSM modem, aj zariadenie umožňujúce napojenie na bezdrôtovú sieť (v našom prípade WiFi sieť). Mobilným operátorom (v prípade GSM siete) a prístupovým bodom (pre bezdrôtovú sieť) by bol automobil napojený na celosvetovú sieť Internet. Cez Internet by bol spojený s expertným systémom, prípadne s iným počítačom určeným výrobcom.

Nakoľko merací systém je určený na diagnostikovanie napájacieho systému automobilu, bude pre bežné diagnostické meranie postačovať svojou rýchlosťou GSM sieť. Nakoľko celý systém by mal pracovať počas bežnej prevádzky, tak sa dá využiť benefit GSM siete, a to je jej veľké pokrytie. WiFi sieť sa bude využívať len pri servisnom zásahu a kontrole meracieho hardwaru, prípadne jeho nastavenia. Taktiež sa môže použiť na zálohu nameraných dát, ktoré sa ukladajú v meracej aparatúre. Takéto WiFi siete by mali mať vybudované špeciálne servisné strediská, kde by bolo možné veľmi rýchlo sa spojiť s Expertným systémom, a na základe porovnania dát doplniť databázu nameraných údajov o tie, ktoré sa poškodili počas prenosu GSM sieťou. Týmto krokom by si expertný systém doplnil chýbajúce údaje, a získal by kompletný prehľad o napájacej sústave daného automobilu.

Ďalším významným prínosom tohto systému je možnosť doladenia parametrov dobíjacej sústavy. Na základe odporúčania Expertného systému (ES) by sa dali navrhnúť lepšie dobíjacie parametre pre daný automobil, prípadne upraviť parametre dobíjacej sústavy na také hodnoty, ktoré lepšie vyhovujú štýlu jazdy a dobe prevádzky auta. Stanovením doby prevádzky automobilu sa dá taktiež zistiť opotrebovanie motora a mnohé ďalšie parametre.

Ako je zrejme z vyššie spomenutých vlastností, celá navrhnutá aparatúra má v sebe veľké množstvo prínosov, ktoré sa dajú implementovať priamo do praxe, a tým uľahčiť diagnostikovanie skrytých chýb v automobile.

III. Realizácia meracieho reťazca

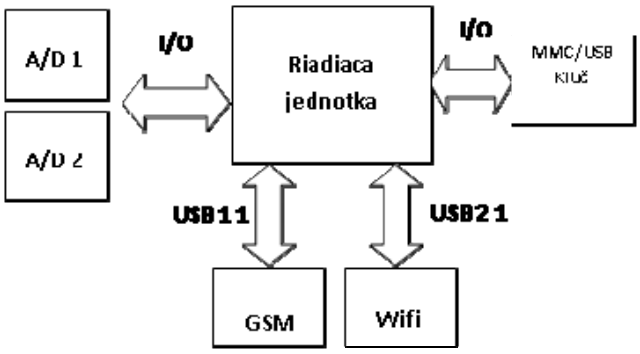
Realizáciu diagnostického systému môžeme rozdeliť do dvoch hlavných častí.

Prvú časť tvorí hardvér, v ktorom je realizovaný celý automatizovaný merací reťazec. Jadrom celého systému je riadiaca jednotka, ktorá riadi pomocné obvody, ktoré merajú napätie a prúd priamo na autobatérii. Tieto namerané hodnoty sa následne, za pomoci analógovo - číslicového prevodníka, digitalizujú a ukladajú do záložnej pamäte. Riadiaci obvod ďalej zabezpečuje pripojenie na Internet za pomoci GSM modemu a Wifi siete. Namerane údaje sú potom priamo posielané cez Internet do expertného systému, kde sa ďalej spracovávajú a vyhodnocujú.

Druhú časť tvorí návrh a realizácia Expertného systému. Tento systém v sebe spája rôzne vyhodnocovacie algoritmy, ktoré navzájom úzko spolupracujú a vyhodnocujú namerane údaje. Jadro celého systému je založené na serverovských aplikáciách, ktoré tvoria hlavnú vyhodnocovaciu časť celého expertného systému.

A. Hardvérová časť diagnostického systému

Bloková schéma hardvérovej časti meracieho reťazca je zostavená z týchto hlavných blokov, ktoré sú znázornené na nasledujúcom obrázku Obr. 2.



Obr. 2. Bloková schéma hardvérovej časti meracieho reťazca.

Jadrom celého meracieho reťazca je riadiaca jednotka, ktorá riadi všetky ostatné obvody a zabezpečuje pripojenie do siete Internet. Meranie priamo zabezpečuje dvojica analógovo – digitálnych prevodníkov. Tieto prevodníky snímajú napätie a prúd priamo na batérii, a následne transformujú namerané analógové veličiny do digitálneho tvaru. Tieto údaje riadiaca jednotka ukladá na USB kľúč, alebo MMC pamäťovú kartu a ďalej ich odosiela pomocou GSM modemu do expertného systému. Wifi pripojenie sa používa na aktualizáciu a nastavenia riadiacej jednotky.

B. Softvérová časť diagnostického systému – Expertný systém

Hlavnou úlohou ES je vyhodnocovať a ukladať namerané údaje. Tieto údaje posiela v pravidelných intervaloch meracia aparatúra vo forme súboru. Na komunikáciu s meracou aparatúrou je použitý FTP server, ktorý je správne nakonfigurovaný a prijaté súbory ukladá na disk do presne vymedzeného diskového priestoru, a následné ich ďalej spracováva.

Keďže hlavným cieľom tohto merania je dlhodobý zber údajov, tak aj množstvo nameraných súborov je veľmi veľké. Na spracovanie takéhoto množstva údajov sa najlepšie javí použiť databázový systém, ktorý je predurčený na spracovávanie a uchovávanie veľkého množstva dát. Tento databázový systém nedokáže bližšie vyhodnocovať namerané údaje, preto je potrebné ho spojiť s nejakým nástrojom, ktorý dokáže namerané údaje vyhodnotiť. Najlepším riešením sa javí použiť PHP skript, ktorý dokáže úzko spolupracovať s databázovým systémom. Veľkou výhodou tohto riešenia je, že namerané údaje dokáže prehľadne zobraziť aj na WWW stránke. Týmto spôsobom dokáže namerané dáta vizualizovať a dokonca, pri správnom nastavení vyhodnocovacích skriptov, aj nájsť chybu, prípadne nejakú anomáliu.

Takže na správne fungovanie ES je potrebná nasledujúca štvorica programov, ktorá tvorí hlavnú kostru tohto systému:

- FTP - server
- MySQL – databázový systém.
- Apache –HTTP server
- PHP – skriptovací jazyk

Všetky tieto programy sú voľne dostupné na Internete, takže na zriadenie a vybudovanie takéhoto ES nie sú potrebné žiadne investície.

IV. VYHODNOCOVANIE NAMERANÝCH ÚDAJOV ES

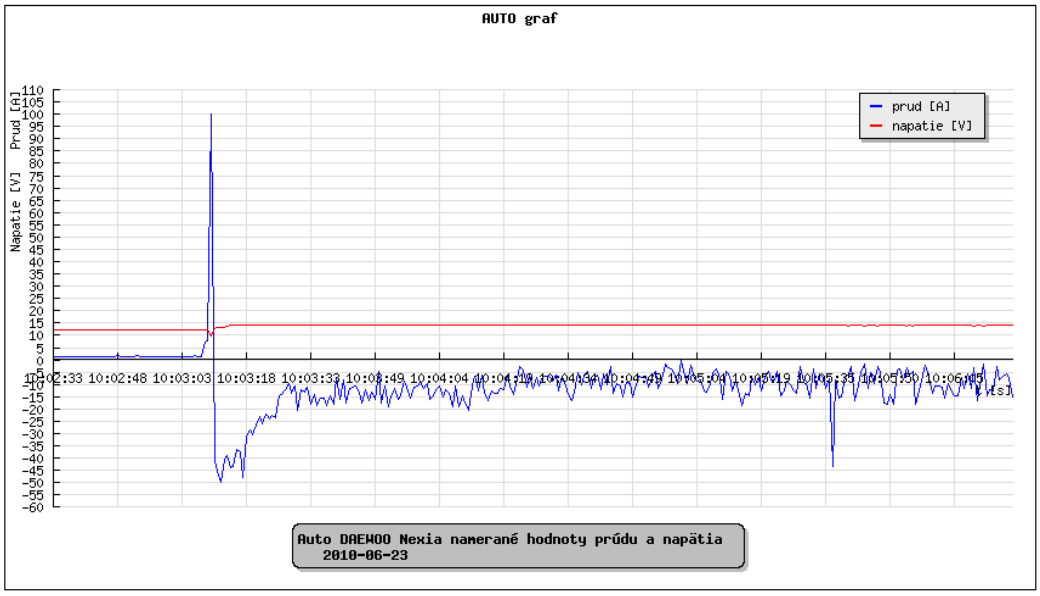
Namerané dáta, ktoré pošle meracia aparatúra sa po prijatí nielen zobrazujú, ale aj vyhodnocujú expertným systémom. ES, tento dokáže namerané dáta uložiť do databázy. V tejto databáze sú dáta uložené vo forme tabuľky, ktorá je zobrazená na Obr. 3. Tieto údaje sú ďalej vyhodnocované, a na základe stanovených kritérií daný systém dokáže vyhľadať určité anomálie a chyby.

			ID:	Napatie	Prud	Datum	Cas	Porucha	Startovanie	Vypnutie
ukaz	ukaz iba prud	ukaz iba napatie	6678913	10.61888	99.973602	2010-06-28	11:19:30		X	
ukaz	ukaz iba prud	ukaz iba napatie	6676707	11.52656	99.973602	2010-06-28	10:42:22		X	
ukaz	ukaz iba prud	ukaz iba napatie	6673059	11.1752	99.973602	2010-06-28	09:40:57		X	
ukaz	ukaz iba prud	ukaz iba napatie	6671338	10.69696	99.973602	2010-06-28	09:11:59		X	
ukaz	ukaz iba prud	ukaz iba napatie	6595261	11.11664	76.793602	2010-06-27	11:11:20		X	
ukaz	ukaz iba prud	ukaz iba napatie	6584171	9.4672	84.211197	2010-06-27	07:54:39		X	
ukaz	ukaz iba prud	ukaz iba napatie	6515130	9.87712	94.947197	2010-06-26	12:12:25		X	
ukaz	ukaz iba prud	ukaz iba napatie	6517779	10.76528	99.973602	2010-06-26	12:57:01		X	
ukaz	ukaz iba prud	ukaz iba napatie	6427807	9.69168	88.896004	2010-06-25	11:32:24		X	
ukaz	ukaz iba prud	ukaz iba napatie	6349355	10.79456	99.973602	2010-06-24	13:01:45		X	
ukaz	ukaz iba prud	ukaz iba napatie	6259331	11.43872	80.502403	2010-06-23	11:36:16		X	
ukaz	ukaz iba prud	ukaz iba napatie	6253813	9.3696	99.973602	2010-06-23	10:03:22		X	

Obr. 3. Základná tabuľka so záznamom ES.

Záznam v tabuľke sa automaticky refrešuje po stanovenom časovom intervale. Takže v tabuľke sú zobrazené stále posledné prijaté údaje z meracej aparatúry. Keď ES nájde určenú nezrovnalosť, pridá špeciálnu značku do tabuľky, aby bolo lepšie identifikovateľne, že sa jedná o nejakú nezrovnalosť. Pod pojmom nezrovnalosť nie je potrebné chápať hneď nejakú chybu. ES systém zaznamenáva informácie aj o štartovaní a vypínaní automobilu. Tabuľka tak obsahuje viacero stĺpcov, aby bolo možné implementovať do nej viacero značiek.

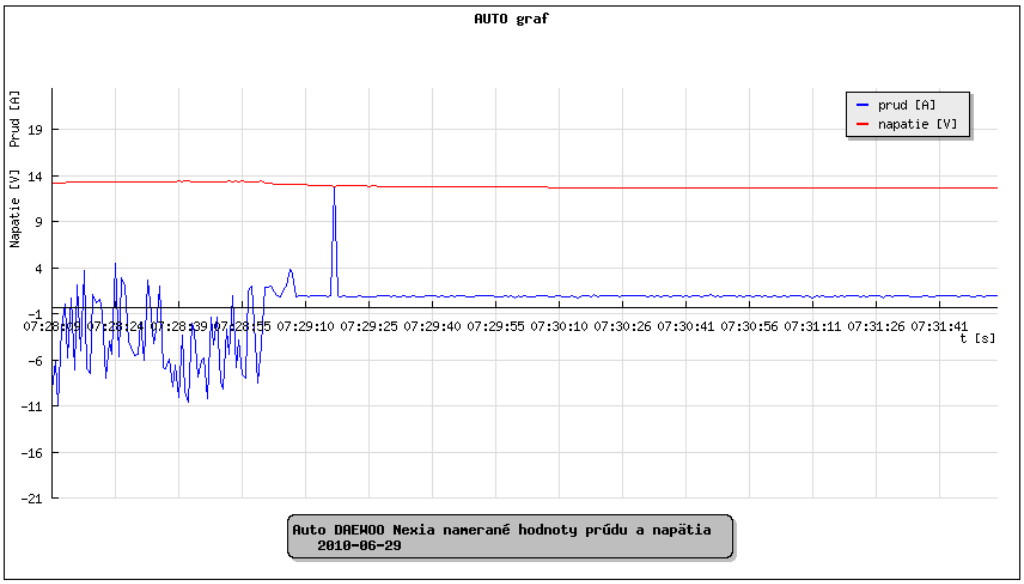
Ako je zrejme z Obr. 3. je v tabuľke v stĺpci „Startovanie“ zaznačená značka, to znamená že ES takto záznam vyhodnocuje ako štart. Nakoľko „štart“ je aj po označení a identifikovaní ES dosť ťažko predstaviteľný len na základe informácií o hodnote napätia a prúdu v danom čase, na lepšie pochopenie udalosti je urobená aj vizualizácia vzniknutého stavu. Takže ES dokáže namerané stavy aj zobraziť. Na obrázku Obr. 4 je zobrazený štart automobilu v grafickej podobe.



Obr. 4. Grafické znázornenie štartu automobilu.

Na Obr. 4. sú na osi y naznačené hodnoty napätia a prúdu, na osi x je informácia o čase. V spodnej časti obrázka je zobrazená informácia o automobile a o dátume, kedy bolo urobené meranie. Na základe zobrazených informácii, si vieme urobiť ucelený obraz o vzniknutom stave.

Ďalší stav, ktorý dokáže identifikovať ES je vypnutie automobilu. Tu je veľmi dôležité ho správne určiť, nakoľko už nie je také jednoznačné ako štart, ale tiež má svoje charakteristické vlastnosti. Vypnutie automobilu je zobrazené na nasledujúcom obrázku Obr. 5.

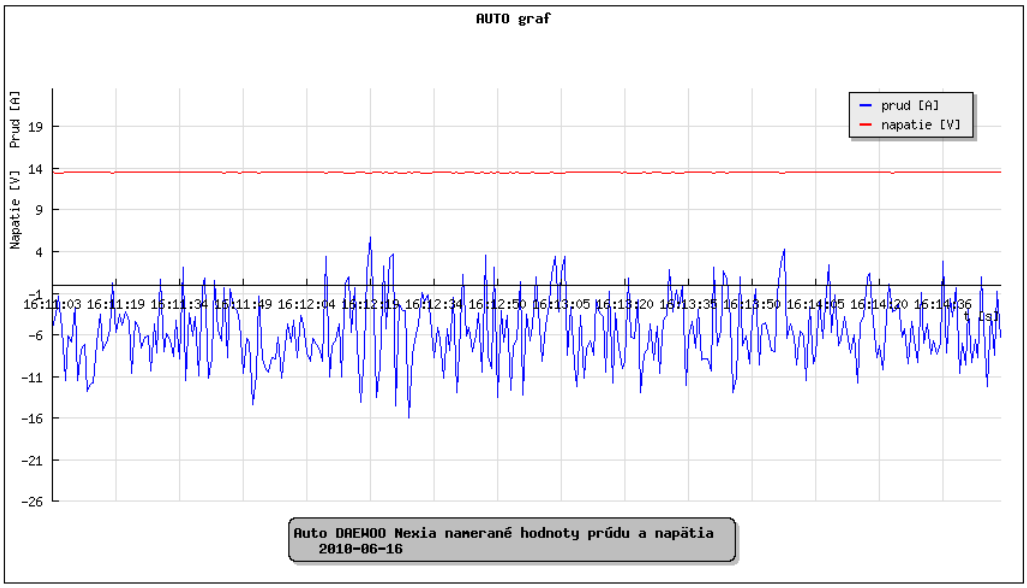


Obr. 5. Grafické znázornenie vypnutia automobilu.

Na prvý pohľad je zrejmé, že na obrázku je vykreslená iná simulácia ako štart auta. Vypnutie má tiež svoje špecifické vlastnosti, jednoznačne je definované vyššou hodnotou napätia a špičkou prúdu, ktorá vzniká pri vypnutí motora, kedy dobíjacia sústava nedokáže pokryť spotrebovanú energiu. ES potom dokáže na základe týchto kritérií identifikovať vypnutie auta.

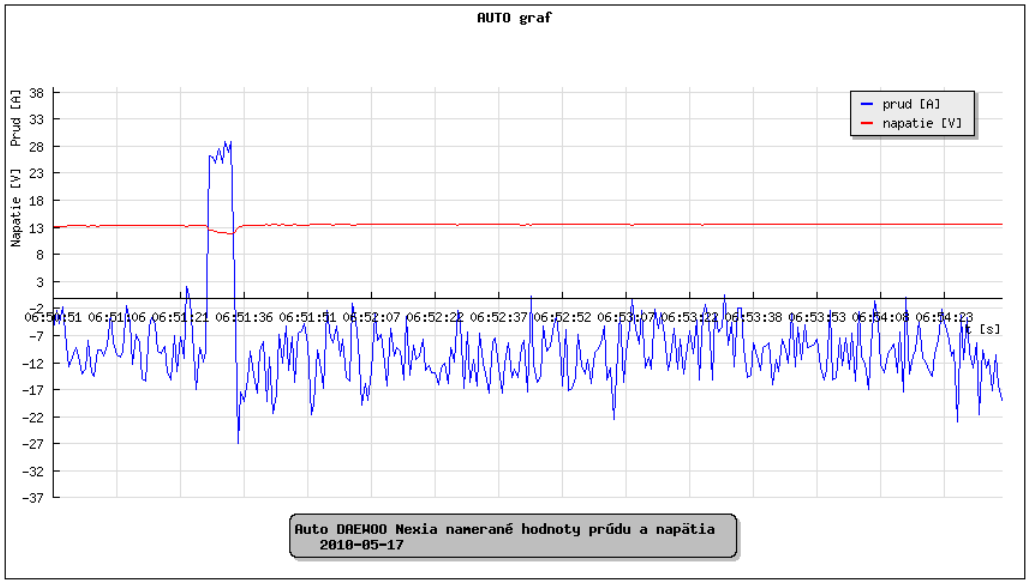
Na Obr. 5. je zobrazené štandardné vypnutie automobilu, tu môžu vznikať chyby, ktoré majú za príčinu znižovanie životnosti autobatérie. Tieto chyby spôsobujú systémy, ktoré sa odpájajú postupne po vypnutí auta, ako napr. chladenie, alarm, rádio a pod. Týmto meraním sa tieto chyby dajú odhaliť a správnym načasovaním vypínania potom odstrániť. V predchádzajúcich častiach bola charakterizovaná a popísaná dvojica normálnych základných prevádzkových stavov, ktoré vznikajú pri prevádzke automobilu. Práve ich nájdením a porovnaním s aktuálne meranými údajmi môžeme eliminovať chyby, ktoré môžu vznikať pri štarte a vypnutí.

V jednoduchosti by sa dalo povedať, že všetky ostatné stavy, mimo základných prevádzkových, sú poruchy, ale nie je to celkom jednoznačné, preto nastavenie kritérií identifikácie porúch bolo pri vytváraní ES systému najzložitejšie. Pre lepšie pochopenie problematiky sú na nasledujúcom obrázku Obr. 6. zobrazené namerané údaje počas behu motora.



Obr. 6. Grafické znázornenie bežnej prevádzky automobilu.

Hodnoty pri behu motora už nie sú také jednoznačné ako je to pri vypnutí alebo pri štarte. Preto je hľadanie chýb veľmi náročné, ale správnym nastavením kritérií ES sa poruchy dajú vyhľadať, pričom sa následne táto porucha označí v tabuľke. Kritérium na hľadanie chýb musí vychádzať z vlastnosti a povahy chyby. Najčastejšie sa jedná o skrat nejakého elektrického spotrebiča alebo o poruchu dobíjacej sústavy. Preto bolo možné určiť podmienky vyhodnocovania poruchového stavu, buď išlo o prudký pokles napätia, alebo o prudký nárast spotrebovaného prúdu. Na základe týchto podmienok dokáže ES určiť poruchu počas behu motora. Na nasledujúcom obrázku Obr. 7. je znázornená porucha auta, pričom táto bola jednoznačne identifikovaná ES.



Obr. 7. Grafické poruchy napájacej sústavy automobilu.

Takže ES dokáže spoľahlivo identifikovať tri základne stavy, ktoré môžu vzniknúť pri používaní automobilu. Jeho ďalšou výhodou je to, že dokáže namerané údaje prehľadávať aj na základe kritérií, ktoré môže zadať obslužný personál, tým sa eliminuje možnosť nezachytenia vzniknutej chyby, ktorá by sa mohla vymykať stanoveným podmienkam a vytvára sa možnosť jeho ďalšieho rozširovania.

Veľkým prínosom tohto systému je to, že dokáže informovať o vzniknutej poruche správcu systému zaslaním mailu na definovanú adresu. Týmto spôsobom vzniknutú chybu dokáže posúdiť kvalifikovaná osoba, a zaujať k nej stanovisko. Chyba objavená systémom je buď periodicky sa opakujúca alebo je to jednorazová chyba. Na základe jej posúdenia môže servis kontaktovať príslušného majiteľa automobilu a vyzvať ho, aby sa dostavil na prehliadku.

V. ZÁVER

Cely článok poskytuje komplexný pohľad na problematiku vzdialeného merania a implementovania nových trendov do tejto oblasti, ktoré sú vhodné pre použitie v automobilovom priemysle. Poskytuje ucelený návod pre zostavenie celého meracieho systému, ktorý spolupracuje z vyhodnocovacím expertným systémom, spolu tvoria komplexný merací a diagnostický systém. Celé riešenie je komplexne jednoduché a ekonomicky nenáročné. Merací systém možno veľmi jednoducho rozširovať, tým sa otvárajú aj otázky prepojenia tohto systému priamo so zbernicou automobilu. Týmto prepojením by sa otvorili nové možnosti merania a automobil by sa natrvalo stal súčasťou siete Internet.

POĎAKOVANIE

Článok bol vypracovaný za podpory grantového slovenského projektu KEGA č 005TUKE-4/2012.

LITERATÚRA

[1] Kováč, D., Kováčová, I., Molnár, J.: Electromagnetic compatibility – measurement (in Slovak), Technical university of Košice, Košice 2008, 1st edition, ISBN 978-80-553-0151-8
[2] Kováč, D., Vince, T., Molnár, J.: Modern Internet Based Production Technology. In: New Trends in Technologies: Devices, Computer, Communication and Industrial Systems, Book edited by: Meng Joo Er, Publisher: SCIYO, November 2010, pp. 145 - 164, ISBN 978-953-307-212-8
[3] Vince T., Kováčová I.: Distance control of mechatronic systems via Internet, In: Acta Electrotechnica et Informatica, 2007, Roč. 7, č. 3 (2007), s. 63-68. - ISSN 1335-8243
[4] Vince, T., Molnár, J., Bučko R.: Real-time regulation systems based on internet - optimization algorithm. In: Transactions of KMOSU. Vol. 4, no. 1 (2010), p. 150-153. ISSN 2072-8263
[5] Tomčíková, I., Molnár, J., Vince, T.: Interaction of magnetic field and tenseness for elastomagnetic sensor of pressure force (in Slovak). In: Elektrorevue, No. 34 (2008), pp. 34-1-34-11
[6] Vince, T., Kováč, D: Remote and regulation of electrotechnical systems via Internet (in Slovak). Publisher: FEI TU Košice 2010, 1st edition, 130 pages, ISBN 978-80-553-0571-4

Vývoj softvéru pre vzdialené meranie variabilných topológií

Tibor VINCE

Katedra teoretickej elektrotechniky a elektrického merania, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská Republika

tibor.vince@tuke.sk

***Abstrakt* —V tomto článku je prezentovaný vyvíjaný softvér pre vzdialené meranie elektrických obvodov s variabilnou topológiou. Merací systém umožňuje fyzicky vytvoriť elektrickú schému na diaľku cez Internet. Užívateľ môže pomocou tohto systému navrhnúť elektrickú schému, v ktorej môže použiť prvky elektrického obvodu z definovanej množiny prvkov. Následne systém navrhnutú schému aj fyzicky zostaví pomocou reálnych elektrických prvkov. Po zostavení obvodu môže užívateľ riadiť napätie na riaditeľných zdrojoch a odčítať hodnoty z meracích prístrojov, ktoré použil v schéme. Systém ako taký je tvorený serverovou, klientskou časťou a špeciálnym laboratórnym stolom s konkrétnymi elektrickými prvkami. Prezentovaný systém môže byť využívaný ako na edukačné, tak aj na vedecké účely merania parametrov elektrických schém. Tento príspevok je zameraný hlavne na softvérovú časť systému.**

***Keywords* — Internet, Vzdialené meranie, Klient-Server, Softvér**

I. Úvod

Jedno z hlavných edukačných zameraní Katedry teoretickej elektrotechnike a elektrického merania je práve výučba teórie o elektrických obvodoch a praktická výučba elektrického merania. Akademické prostredie je silne ovplyvnené vývojom a inováciou v tejto oblasti a taktiež výstupmi špecifických výskumných projektov. Pri našom snažení o udržaní kroku s modernými systémami, inováciami a technológiami, narážame na mnohé problémy v procese výučby. Jedným z nich je problém časového limitu – nedostatok rozsahu hodín a nedostatok priestorov pre výučbu. Typická situácia je, keď vyučujúci na cvičení má na starosti 12 až 15 študentov. Bežná úloha študentov je poskladať elektrický obvod podľa aktuálne preberanej látky a vykonať sériu meraní konkrétnych parametrov obvodu. Z dôvodu nedostatku času, priestoru a učiteľov, prakticky len niekoľkí študenti majú možnosť poskladať obvod. Väčšinou sa jedná o študentov, ktorí majú praktické skúsenosti už zo strednej školy. Väčšina študentov je len v úlohe pozorovateľov, ktorí nemajú možnosť aktívne prispieť k priebehu merania. Kapacita laboratórií a počet vyučujúcich neumožňujú, aby si každý z niekoľko stoviek študentov vytvoril sám obvod a aplikoval vlastné merania a tým získal praktické skúsenosti.

Problémom je taktiež relatívne veľká náročnosť zaradenia predmetov tohto typu predmetov do dištančného vzdelávania, hlavne vzdelávania typu e-Learning. V dnešnej dobe je pomerne bežné, že sa Internet využíva v procese výučby založenej na sprístupňovaní dokumentov v elektronickej forme. V minulosti bol pojem e-Learning používaný pre učebné metódy využívajúce elektronické pomôcky. Dnes sa však stal synonymum dištančného vzdelávania cez Internet[1]. Internetové spojenie umožňuje distribuovať využitie zdrojov ako hardvér, softvér, dáta, znalosti o mnohé ďalšie celosvetovo. Umožňuje tak využívanie týchto zdrojov študentom alebo pracujúcim v ich domácom prostredí s minimálnymi nákladmi. Prvé pokusy o využívanie e-Learning-u bolo založené na systéme samo štúdia.

V rámci e-Learning-u nie je problém v problematike elektrického merania vytvoriť teoretické podklady pre učebnú látku formou textov, hypertextov, obrázkov, popisom konkrétnych cvičení, virtuálnych nástrojov a simulácií. Avšak zvyšujúci sa počet študentov, cena zariadení, komplikovanosť ich nákupu a súčasne znižujúci sa rozpočet laboratórií je stále väčšou hnacou silou a zdôraznením o potrebe sprístupnenia týchto laboratórnych prístrojov na Internet. Ideálnym stavom by bolo, aby študenti pomocou takéhoto vzdialeného prístupu v spojení s e-Learning-om mohli získať rovnaké skúsenosti, ako by dané cvičenie praktizovali v laboratóriu. Existuje len malé množstvo aplikácií, ktoré reálne umožňujú prístup užívateľom priamo k

meracím zariadeniam. Väčšinou sa jedná o systémy založené na koncepte Klient - Server. Ten funguje na princípe, že študent má na svojom počítači nainštalovanú klientsku časť programu a v laboratóriu je spustená serverová časť systému, ktorá komunikuje s konkrétnymi meracími prístrojmi. V súčasnosti je celkom bežné vzdialené meranie, kde je pre prenos dát využívaný práve Internet. Niektoré aplikácie umožňujú aj interaktivitu. Väčšinou sa jedná o aplikácie, ktoré umožňujú zapnúť alebo vypnúť merací reťazec, prípadne nastaviť parametre meracieho reťazca [2][3]. Z momentálne existujúcich vzdialených meracích reťazcov niektoré ešte umožňujú prepínať medzi vopred definovanými topológiami.

Jednou zo snáh na riešenie vyššie uvedených problémov je vyvíjaný systém VMLab, ktorého úlohou je umožniť študentom navrhnúť obvod s ľubovoľnou topológiou, navrhnutý systém fyzicky poskladať na diaľku a namerané údaje odoslať študentovi. [4]

II. VMLAB

VMLab znamená skratku pomenovania **V**zdialené **M**eracie **Lab**oratórium. Jedná sa o unikátny systém vyvíjaný na Katedre teoretickej elektrotechniky a elektrických meraní (KTEEM) FEI TU Košice. Systém umožňuje z ktoréhokoľvek miesta s prístupom na Internet vytvoriť navrhnúť elektrický obvod z vopred definovanej množiny elektrických prvkov. Následne je navrhnutý obvod aj fyzicky poskladaný v laboratóriu z reálnych komponentov. Niektoré z elektrických komponentov môžu byť riadené (nastaviteľné). V tom prípade je možné počas samotného merania nastavovať parametre daného komponentu. Príkladom môžu byť napäťové zdroje, kde užívateľ môže meniť hodnotu napätia na použitých zdrojoch podľa potreby. Všetky namerané údaje sú potom prenesené aktuálnemu užívateľovi, ktorý vytvoril daný elektrický obvod.

VMLab je postavený na báze klient – server aplikácie. Schéma samotného systému je zobrazená na obrázku Fig. 1.

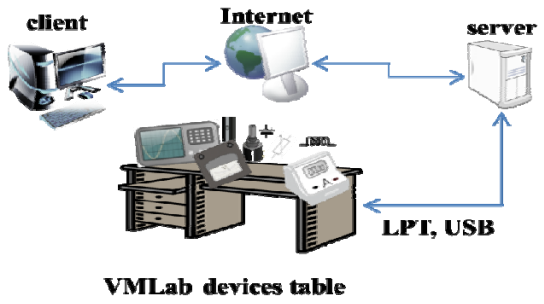


Fig. 1 Schéma VMLAb.

VMLab je možné rozdeliť na nasledujúce časti: klient, server a VMLab stôl. VMLab stôl pozostáva z konkrétnych elektrických komponentov. Elektrické komponenty môžu byť pasívne prvky ako odpory, cievky, kondenzátory, potom meracie prístroje napr. voltmeter, ampérmeter, wattmeter a taktiež zdroje elektrického napätia. Všetky komponenty sú pripojené do špeciálneho hardvéru, ktorý vytvára prepojenia medzi jednotlivými komponentmi podľa návrhu elektrického obvodu. Tento špeciálny hardvér je súčasťou systému VMLab a bol taktiež vyvíjaný na katedre KTEEM je ovládaný pomocou servera. VMLab stôl je zobrazený na obrázku Fig. 2.

Server prijíma dáta od klienta. Podľa prijatých dát potom odosiela inštrukcie VMLab stolu pre vytvorenie spojenia konkrétnych elektrických komponentov. Taktiež však komunikuje s konkrétnymi meracími zariadeniami a s riaditeľnými zdrojmi. Namerané údaje potom posiela klientovi, poprípade nastavuje hodnotu riaditeľných zdrojov podľa požiadaviek klienta.

Klientska časť je zobrazená na obrázku Fig 2. Klientska časť beží na počítači konkrétneho užívateľa. Užívateľ pomocou tejto aplikácie vytvorí návrh elektrickej schémy a následne návrh odošle na server. Taktiež prijíma od servera informácie o aktuálne nameraných hodnotách a užívateľovi umožňuje nastavovať hodnotu riaditeľných zdrojov. Najväčšou devízou systému VMLab je, že nie je založený na aplikácii niektorej z vopred definovaných topológií, ale vytvára práve takú topológiu, ktorú navrhni užívateľ. Existujúci prototyp systému VMLab má momentálne obmedzenia 64 pólov a 16 uzlov. To znamená, že do VMLab stola je možné pripojiť maximálne 32 dipólov alebo 16 štvorpólov, prípadne primeranú kombináciu elektrických komponentov s maximálnym súčtom pólov 64. Druhé obmedzenie pri vytváraní elektrických obvodov je maximálny počet uzlov: 16. Avšak uvedené obmedzenia VMLab

systému sú obmedzenia existujúceho prototypu. Nie sú to obmedzenia systému VMLab ako takého. V prípade potreby je možné vytvoriť VMLab systém s väčšou kapacitou pólov alebo uzlov.



Fig. 2 VMLab stôl a server.

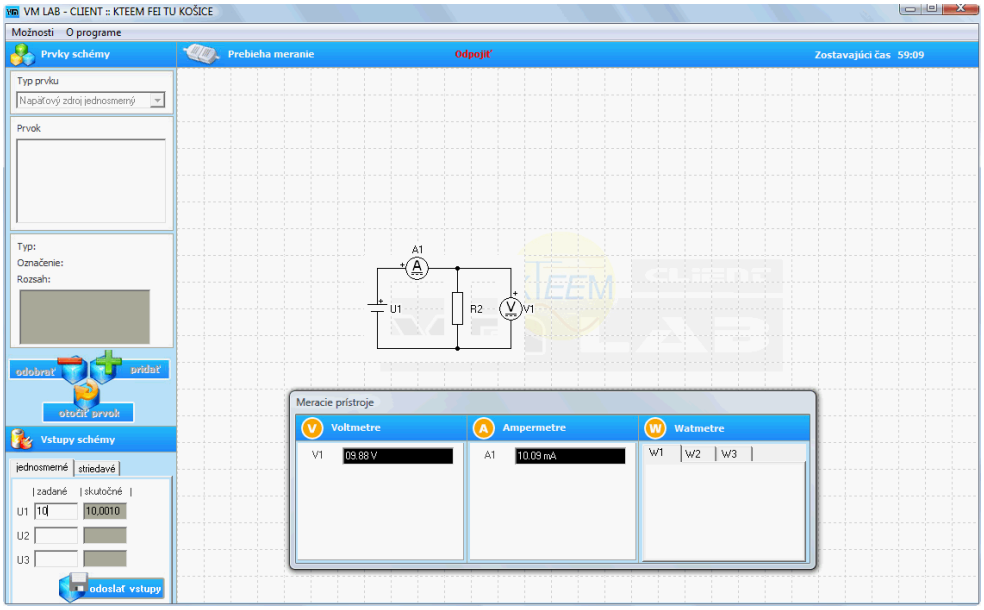


Fig. 3 VMLab klient.

III. Popis činnosti

Po spustení klientskej aplikácie užívateľom, ako prvý krok je spojenie so serverom. Od servera získa informácie o aktuálne pripojených elektronických prvkoch. VMLab umožňuje meniť množinu elektrických komponentov napojených na VMLab stôl. Po získaní aktuálneho zoznamu môže užívateľ začať navrhovať elektrický obvod. V návrhu obvodu mu program umožní použiť len tie komponenty, ktoré sa nachádzajú na aktuálnom zozname komponentov. Navrhnutú schému môže samozrejme uložiť do súboru, prípadne ju načítať zo súboru.

Po vytvorení schémy sa môže užívateľ pripojiť k VMLab stolu. V jednej chvíli môže VMLab stôl používať maximálne jeden užívateľ. Preto, ak je VMLab stôl obsadený, musí čakať na jeho uvoľnenie. Pre pripojenie k VMLab stolu sa musí užívateľ autentifikovať menom a heslom. Po úspešnej autentifikácii klientsky softvér odošle serveru parametre navrhnutého obvodu. Podľa odoslaných parametrov server nakonfiguruje VMLab stôl tak, aby vznikli prepojenia medzi jednotlivými komponentmi podľa navrhovanej schémy. Táto konfigurácia je odoslaná VMLab stolu cez LPT port. Po vytvorení obvodu server začne komunikovať s meracími a riaditeľnými komponentmi použitými v elektrickom obvode. Načíta údaje o reálnych hodnotách jednotlivých meracích prístrojov a riaditeľných zdrojov. Komunikácia s týmito prístrojmi je realizovaná pomocou rozhrania USB a RS232. Takto získané údaje sa potom odošlú klientovi. Klient počas procesu merania môže nastavovať hodnoty zdrojov, ktoré používa v schéme. Akékoľvek zmena žiadaného napätia je prenášaná znova na server

a následne odoslaná konkrétnemu zdroj napätia. Zmena napätia sa udeje za niekoľko sekúnd od odoslania požiadavky. Počas merania si užívateľ môže ukladať jednotlivé namerané hodnoty. Po ukončení merania sa elektrický obvod vytvorený na VMLab stole rozpojí. Užívateľ si následne môže vygenerovať dokumentáciu o meraní, v ktorej je zaznamenaná schéma, zoznam použitých elektrických komponentov a tabuľka nameraných hodnôt.

IV. ZÁVER

Stúpajúci počet študentov, ceny prístrojov a znižujúci sa rozpočet laboratórií vedie k problémom s časovými, finančnými a ľudskými zdrojmi. Jedno z najefektívnejších riešení je možnosť vytvorenie laboratória pre automatizované vzdialené meranie. V vyučovacom procese sa môžu študenti pomocou VMLab systému pripojiť do laboratória a aplikovať proces merania s navrhnutou schémou. Najväčšou výhodou VMLab systému je, že VMLab výsledky merania nie sú len počítačovou simuláciou, ale sú fyzicky odmeranými dátami. Užívatelia sa môžu pripojiť k VMLab systému 24 hodín denne, 7 dní v týždni z akéhokoľvek miesta s prístupom na Internet. V rámci zariadení vo VMLab stolu môžu byť použité aj špičkové meracie zariadenia s veľmi vysokou presnosťou. VMLab by mohli použiť nie len študenti, ale by mohol byť otvorený širšej komunite výskumníkov. Systém VMLab je predmetom patentového konania.

POĎAKOVANIE

Tento príspevok bol vytvorený s podporou Slovenského grantového projektu KEGA č. 005TUKE-4/2012

Podporujeme výskumné aktivity na Slovensku/ Projekt je spolufinancovaný zo zdrojov EÚ. Tento článok bol vypracovaný v rámci projektu "Centrum excelentnosti integrovaného výskumu a využitia progresívnych materiálov a technológií v oblasti automobilovej elektroniky", ITMS 26220120055.

REFERENCES

[1] G. Canfora, P. Daponte, S. Rapuano: Remotely accessible laboratory for electronic measurement teaching, Computer Standards & Interfaces, Volume 26, Issue 6, October 2004, Pages 489-499
[2] Peretto, L., Rapuano, S., Riccio, M., Bonatti, D: Distance learning of electronic measurements by means of measurement set-up models, Measurement. Vol. 41, no. 3, pp. 274-283. Apr. 2008
[3] Arpaia, P., Baccigalupi, A., Cennamo, F., Daponte, P.: A Measurement Laboratory on Geographic Network for Remote Test Experiments, IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT, VOL. 49, NO. 5, OCTOBER 2000
[4] Kováč, D., Vince, T. , Molnár, J.: Modern Internet Based Production Technology. In: New Trends in Technologies, SCIYO Publisher, Croatia, 2010, 20 pages,ISBN 978-953-307-212-8
[5] Molnár, J., Kováčová, I.: Distance remote measurement of magnetic field. In: Acta Electrotechnica et Informatica, 2007, Vol.7, No.4, pp. 52-55, ISSN 1335-8243.
[6] Molnár, J., Kováčová, I.: Distance remote measurement of magnetic field. In: Acta Electrotechnica et Informatica, 2007, Vol.7, No.4, pp. 52-55, ISSN 1335-8243.

The crossing number of $G \square P_n$ for the specific graph G of order six

Marián KLEŠČ, Daniela KRAVECOVÁ, Matúš VALO

Dept. of Mathematics and Theoretical Informatics, FEI TU of Košice, Slovak Republic

marian.klesc@tuke.sk, daniela.kravecova@tuke.sk, matus.valo@tuke.sk

Abstract **– The problem of determination and reducing the number of crossings in a graph is studied not only in the theory of graphs, but also in computer science. It is well known that the problem of determination of the crossing numbers of graphs is NP–complete. The exact value of the crossing number is known only for few classes of graphs. Cartesian product of two graphs is one of the first classes of graphs for which the crossing number has been studied. The crossing numbers of Cartesian products of paths P_n, cycles C_n or stars S_n with all graphs of order at most four are known. For the path P_n of length n, the crossing numbers of Cartesian products $G \square P_n$ for all connected graphs G on five vertices are also known. Let G be a graph created from C_4 by adding vertices w_1, w_2 and edges v_1w_1, v_2w_2, where v_1, v_2 are two non-adjacent vertices in C_4. In this paper, the crossing number of the Cartesian product of path P_n and G is studied. We have proved that the Cartesian product of such graphs is $2(n-1)$.**

Keywords **– graph, Cartesian product, crossing number, path**

I. Introduction

Introduction of the VSLI technology revolutionized circuit design and had a big impact on parallel computing. Many researchers have focused on optimizing the VLSI circuit layout. VSLI theory has expanded in many fruitful and interesting directions. One of the branches is layout theory which studied the efficiency of embedding the graphs in the plane according to VSLI rules. A lot of research aiming at efficient use of the new technologies has been done and further investigation are in progress. One of the major tasks is minimizing the number of wire crossings in a circuit, as this greatly reduces the chance of cross-talk in long crossing wires carrying the same signal and also allows for faster operation and less power dissipation. Clark Thompson introduced a simple, graph-theoretic model for very large scale integration circuitry [1]. In Thompson's model, a circuit is a graph whose vertices corresponds to active circuit elements and whose edges correspond to wires. The question of finding the minimal number of crossing wires can be abstracted to a graph theoretical problem of determining the minimal number of edge crossings in a drawing of a given graph. A *finite, undirected, simple graph* $G = (V, E)$ is an ordered pair consisting of the finite set V of vertices and finite set $E \subseteq \{\{u, v\} | u \in V, v \in V, u \neq v\}$ of edges. The vertex set and edge set of the graph G will be denoted by $V(G)$ and $E(G)$, respectively. We will abuse terminology and call finite, undirected, simple graph as *graph*. The edge we will denote by uv instead of $\{u, v\}$. A *drawing* D of a graph G on the plane is a one-to-one mapping of each vertex to a point in $\mathbb{R}^2$, and each edge to a simple open curve between corresponding two vertices. For simplicity, we assume that in a drawing

a) no edge passes through any vertex other than its end-points,
b) no two edges touch each other and
c) no three edges cross at the same point.

A *crossing* is a common point of two curves, different from their endpoints. The crossing number $\mathrm{cr}(G)$ of a graph G is the minimum possible number of crossings in an arbitrary drawing of the graph G. A *good drawing* is a drawing such that:

a) no edge crosses itself,
b) no two edges incident with the same vertex cross each other,
c) no two edges cross more than once.

It can be easily proved that every drawing with minimum number of crossings is a good drawing, see Figure 1. Pioneering papers about the crossing numbers were published in the second half of the twentieth century. Turán formulated *The brick factory problem* that arose during World

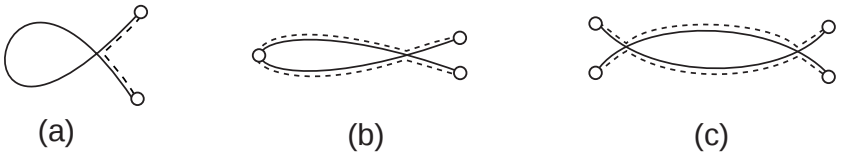


Fig. 1 Elimination of crossings in the good drawing.

War II., which was published in the article [2]. It is a graph problem where kilns and storage yards are vertices and the rails between kilns and storage yards are edges in the graph. Turán asked if the number of crossings of the rails had been minimized. He realized after several days that the actual situation could be improved, but the exact solution of the general problem with m kilns and n storage yards ($K_{m,n}$) seemed to be very difficult. Using the construction of suitable diagram of the graph $K_{m,n}$, Zarankiewicz [3] published the solution that the crossing number of the complete bipartite graph $K_{m,n}$ is

$$\left\lfloor \frac{m}{2} \right\rfloor \left\lfloor \frac{m-1}{2} \right\rfloor \left\lfloor \frac{n}{2} \right\rfloor \left\lfloor \frac{n-1}{2} \right\rfloor .$$

The given problem has not been generally proved, but there are crossing numbers values known for some special cases. The crossing number of $K_{m,n}$ is proved for every n and for $1 \leq m \leq 6$ [4]. Woodall published that for $7 \leq m \leq 8$ and $7 \leq n \leq 10$, the crossing number of $K_{m,n}$ is equal to Zarankiewicz number, too [5].

The complete graph K_n is included among the classes of graphs which high regularity helps to state the crossing number of this graphs. Like the graphs $K_{m,n}$, the problem of determining the crossing number of K_n is not generally resolved. The crossing number of K_n is known for $n \leq 12$ [6], [7].

In addition, the crossing numbers of multipartite graphs are studied, too. Other families of the graphs which have rich symmetry are generalized Petersen graphs and circulant graphs whose crossing numbers were studied by several authors.

Further results are known on the Cartesian product of two graphs. The Cartesian product of the graphs is a class that is the most known regarding to the crossing number. The *Cartesian product* $G \square H$ of the graphs G and H is a graph with the vertex set $V(G \square H) = V(G) \times V(H)$ and the edge set $E(G \square H) = \{\{(x_g, x_h), (y_g, y_h)\} | x_g y_g \in E(G) \text{ and } x_h = y_h, \text{ or } x_h y_h \in E(H) \text{ and } x_g = y_g\}$. Cartesian product of two graphs is another class of graphs which is characterized by great regularity but specifying the Cartesian product of two infinite graphs is not simple everytime. Let C_n be the cycle of length n, P_n be the path of length n, respectively, and let S_n denote the complete bipartite graph $K_{1,n}$. By using a suitable diagram it is easy to see that $cr(P_m \square C_n) = 0$ for $m \geq 1, n \geq 3$. Bokal proved that

$$cr(S_m \square P_n) = (n-1) \left\lfloor \frac{m}{2} \right\rfloor \left\lfloor \frac{m-1}{2} \right\rfloor$$

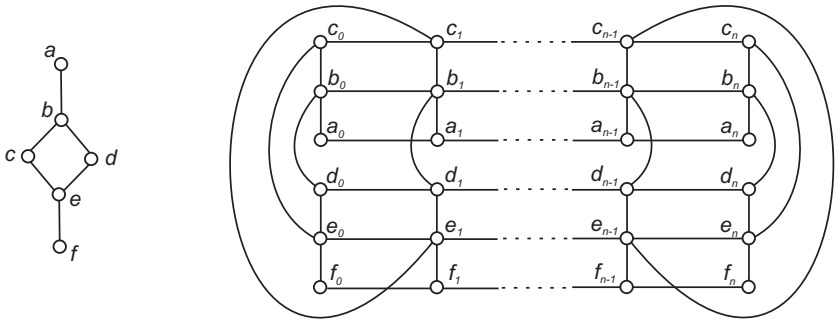
for $m, n \geq 1$ [8]. Among the first works on the Cartesian product $C_m \square C_n$ is included the work of authors Harary, Kainen, Schwenk [9]. They proved that $cr(C_3 \square C_3) = 3$ and conjectured $cr(C_m \square C_n) = (m-2)n$ for $3 \leq m \leq n$. This hypothesis has been proved for $m \leq 7$ by other authors, but it has not yet been proved for $3 \leq m \leq n$ in general.

Several authors were researching the crossing numbers of the Cartesian products $G \square C_n$, $G \square P_n$, $G \square S_n$ for some specific graphs G. If G is a connected graph on two or three vertices the crossing numbers are present in the previous works. Beineke, Ringeisen, Jendroľ and Ščerbová determined the crossing numbers of the Cartesian products of four-vertex graphs with C_n [10], [11]. Klešč established the crossing numbers of the Cartesian products of P_n and S_n with four-vertex graphs in [12]. In [13] and [14] there were published the values for crossing numbers of $G \square P_n$, if G is a connected 5-vertex graph. The crossing number of the Cartesian products $G \square C_n, G \square P_n, G \square S_n$ with 6-vertex graphs is known only for a few graphs.

We denote the number of crossings of a graph G in a drawing D by $\mathrm{cr}_D(G)$. Let G_i and G_j be edge–disjoint subgraphs of a graph G. Then we denote by $\mathrm{cr}_D(G_i, G_j)$ the number of crossings between the edges of G_i and the edges of G_j in the drawing D.

II. The crossing number of $G \square P_n$

In this section we will present exact crossing number of the graph $G \square P_n$. As mentioned above, G is a graph with 6 vertices and 6 edges created from C_4 by adding vertices w_1, w_2 and edges $v_1 w_1$, $v_2 w_2$, where v_1 and v_2 are two non-adjacent vertices in C_4, see Figure 2 (a). We assume



(a) (b)

Fig. 2 The drawing of G and $G \square P_n$

$n \geq 1$ and find it convenient to consider the graph $G \square P_n$ in the following way. It has $6(n+1)$ vertices, which we denote x_i for $x = a, b, c, d, e, f$ and $i = 0, 1, \ldots, n$. The edges belong to $n+1$ copies of the subgraphs isomorphic to G and six paths of length n, see Figure 2 (b). The former one we will denote by $G^i, i = 0, 1, \ldots, n$. For $i = 1, 2, \ldots, n$, let H^i denote the subgraph of $G \square P_n$ containing the vertices of G^{i-1} and G^i and the six edges joining G^{i-1} to G^i. Let $Q^i, i = 1, 2, \ldots, n-1$, denote the subgraph of $G \square P_n$ induced by $V(G^{i-1}) \cup V(G^i) \cup V(G^{i+1})$, alternatively $Q^i = G^{i-1} \cup H^i \cup G^i \cup H^{i+1} \cup G^{i+1}$. Let us denote by Q_b^i or Q_e^i the graph created from Q^i by removing the vertices e_j, f_j or a_j, b_j for $j = i-1, i, i+1$, respectively. It is easy to see that both subgraphs Q_b^i and Q_e^i are isomorphic to $S_3 \square P_2$ and hence $\mathrm{cr}(Q_b^i) = \mathrm{cr}(Q_e^i) = 1$ [8].

Lemma 1: $\mathrm{cr}(G \square P_2) = 2$

Proof: It can be easily seen from Figure 3 that the inequality $\mathrm{cr}(G \square P_2) \leq 2$ holds. By way

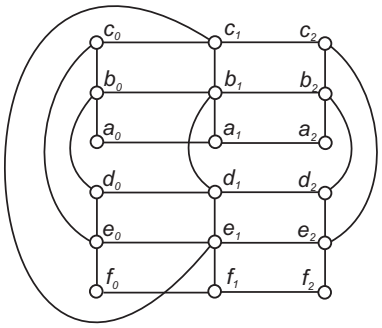


Fig. 3 The drawing of the graph $G \square P_2$.

of contradiction suppose that $\mathrm{cr}(G \square P_2) < 2$ and let D be a good drawing of the graph $G \square P_2$ with less than two crossings. From the definitions of the Cartesian product and the subgraph Q^i, it follows that $G \square P_2 = Q^1$. The subgraph Q^1 contains two aforementioned subgraphs Q_b^1 and Q_e^1, both having crossing number equal to 1. This implies that $\mathrm{cr}(G \square P_2) \geq 1$. If no two common edges of the graphs Q_b^1 and Q_e^1 cross each other, then there are at least 2 crossings in $D(Q^1)$, which contradicts the assumption. Otherwise at least two common edges of the graphs Q_b^1 and Q_e^1 cross each other, which follows that the path $c_{i-1}c_ic_{i+1}$ crosses the path $d_{i-1}d_id_{i+1}$. If this crossing is the only one, the removing one of the crossed edges should create drawing with no crossing. But the removing of an arbitrary edge from the edges $c_{i-1}c_i, c_ic_{i+1}, d_{i-1}d_i$, and d_id_{i+1} leads to the subgraph containing a subdivision of $K_{3,3}$. This contradicts the assumption, because $\mathrm{cr}(K_{3,3}) = 1$. ■

Lemma 2: If D is a good drawing of the graph $G \square P_n, n \geq 3$, in which every subgraph $G^i, i = 0, 1, \ldots, n$, has at most one crossing on its edges, then in D there are at least $2(n-1)$ crossings.

Proof: Consider the good drawing D of $G \square P_n$ assumed in Lemma 2. Let us denote by $C_4^i(G)$ the cycle $b_ic_ie_id_ib_i$ of the graph G^i. We will show at the beginning that from the assumptions of Lemma 2 implies that $\mathrm{cr}_D(C_4^i(G), G^j) = 0$ holds for every $i, j = 0, 1, \ldots, n$, such that $|i-j| = 1$. It is clear that $\mathrm{cr}_D(C_4^i(G), G^j) \leq 1$. Moreover, $\mathrm{cr}_D(C_4^i(G), C_4^j(G)) = 0$, otherwise two disjoint cycles $C_4^i(G)$ and $C_4^j(G)$ cross each other at least twice. Hence, if

$\mathrm{cr}_D(C_4^i(G), G^j) = 1$, the cycle $C_4^j(G)$ is crossed by the edge a_ib_i or by the edge e_if_i. Without loss of generality suppose that $\mathrm{cr}_D(G^{i-1}, C_4^i(G)) = 1$ and that the edge $a_{i-1}b_{i-1}$ crosses $C_4^i(G)$. Then the edges of $C_4^i(G)$ do not cross each other and $C_4^i(G)$ separates the vertex a_{i-1} from the spare vertices of the graph G^{i-1}. Suppose that the vertex a_{i-1} is inside the circle and the spare vertices outside the circle $C_4^i(G)$. Then the vertices f_{i-1} and f_i are outside the circle and the vertex a_i is inside $C_4^i(G)$. Otherwise at least one of the edges $a_{i-1}a_i$ and $f_{i-1}f_i$ crosses $C_4^i(G)$, a contradiction. Hence, for $i \neq n$, the path $a_ia_{i+1}b_{i+1}c_{i+1}e_{i+1}f_{i+1}f_i$ must cross $C_4^i(G)$. For $i \neq n$, the path $a_{i-1}a_{i-2}b_{i-2}c_{i-2}e_{i-2}f_{i-2}f_i$ must cross $C_4^i(G)$. Both cases are in contradiction with the assumption. Thus, $\mathrm{cr}_D(C_4^i(G), G^j) = 0$ for every $i, j = 0, 1, \ldots, n$ with $|i - j| = 1$.

Let us consider the following types of crossings in Q^i:

(1) a crossing between two edges in G^i,
(2) a crossing of an edge in G^i with an edge in $H^{i+1} \cup H^i$,
(3) a crossing of an edge in $G^{i-1} \cup H^i$ with an edge in $G^{i+1} \cup H^{i+1}$
(4) a crossing of an edge in G^i with an edge in $G^{i-1} \cup G^{i+1}$.

It is readily seen that every crossing of types (1), (2) and (3) appears in a good drawing of the graph $G \,\square\, P_n$ only on the edges of the subgraph Q^i. For $i \in \{2, 3, \ldots, n-1\}$, a crossing of type (4) in Q^i between an edge of G^{i-1} and an edge of G^i appears only in Q^{i-1} as a crossing of type (4). For $i \in \{1, 2, \ldots, n-2\}$, a crossing of type (4) in Q^i between an edge of G^i and an edge of G^{i+1} appears only in Q^{i+1} as a crossing of type (4).

In a good drawing of $G \,\square\, P_n$, we define the *force* $f(Q^i)$ of Q^i in the following way: every crossing of type (1), (2) and (3) contributes the value 1 to $f(Q^i)$ and every crossing of type (4) contributes the value $\frac{1}{2}$. The total force of the drawing is the sum of $f(Q^i)$. As every crossing of type (1), (2) or (3) is counted only once and every crossing of type (4) is counted at most twice and no other crossing contributes to the total force of the drawing, the number of crossings in the drawing is not less than the total force of the drawing. So, the aim of this proof is to show that if every of the subgraphs $G^i, i = 0, 1, \ldots, n$, has at most one crossing on its edges, then $f(Q^i) \geq 2$ for all $i = 1, 2, \ldots, n-1$.

Let us denote by G_b^i (G_e^i) the subgraph of G^i induced by the vertices a_i, b_i, c_i, d_i (c_i, d_i, e_i, f_i). It can be easily seen that both G_b^i and G_e^i are isomorphic with the star S_3. Let H_b^i (H_e^i) denote the subgraph of Q^i induced by the edges connecting G_b^{i-1} with G_b^i (G_e^{i-1} with G_e^i). The inequalities $\mathrm{cr}_D(Q_b^i) \geq 1$ and $\mathrm{cr}_D(Q_e^i) \geq 1$ hold since both subgraphs Q_b^i and Q_e^i contain a subdivision of $K_{3,3}$ as a subgraph. As no two adjacent edges cross each other in a good drawing of $K_{3,3}$, the forced crossing in $D(Q_b^i)$ is only one of the following types:

(a) a crossing of an edge in $G_b^{i-1} \cup H_b^i$ with an edge in $G_b^{i+1} \cup H_b^{i+1}$,
(b) a crossing of an edge in G_b^i with an edge in $H_b^i \cup H_b^{i+1}$,
(c) a crossing of an edge in G_b^i with an edge in G_b^{i-1} or G_b^{i+1}.

Every crossing of the type (a) and (b) contributes the value 1 to $f(Q^i)$, and every crossing of the type (c) contributes the value $\frac{1}{2}$ to $f(Q^i)$. The same situation is in $D(Q_e^i)$. If the forced crossings in $D(Q_b^i)$ and $D(Q_e^i)$ are different and both contribute the value 1 to $f(Q^i)$, then $f(Q^i) \geq 2$, and we are done.

Let us consider that the forced crossings in $D(Q_b^i)$ and $D(Q_e^i)$ are identical. Then either the edge $c_{i-1}c_i$ crosses d_id_{i+1} or the edge $d_{i-1}d_i$ crosses c_ic_{i+1}. Without loss of generality, assume that the edges $c_{i-1}c_i$ and d_id_{i+1} cross each other in $D(Q^i)$. Then, in $D(Q^i)$, the vertex-disjoint cycles $c_ic_{i-1}e_{i-1}e_ic_i$ and $d_id_{i+1}b_{i+1}b_id_i$ cross each other at least twice. As $\mathrm{cr}_D(C_4^{i-1}(G), G^i) = 0$ and $\mathrm{cr}_D(C_4^i(G), G^{i+1}) = 0$, the edge $c_{i-1}e_{i-1}$ does not cross b_id_i and the edge c_ie_i does not cross $b_{i+1}d_{i+1}$. Hence, both necessary crossings between the 4-cycles $c_ic_{i-1}e_{i-1}e_ic_i$ and $d_id_{i+1}b_{i+1}b_id_i$ contribute at least 1 to $f(Q^i)$ and $f(Q^i) \geq 2$ in the case of common crossing in $D(Q_b^i)$ and $D(Q_e^i)$.

It implies from the paragraph above that if the forced crossings in $D(Q_b^i)$ and in $D(Q_e^i)$ are not identical, then the inequality $f(Q^i) \geq 1,5$ holds. The contribution with the value $1,5$ is only in the case when one of the considered crossings is a crossing of type (c). Such a crossing can appear only once per G^i according to the assumption of the lemma.

Assume, without loss of generality, that G_b^{i-1} crosses G_b^i in $D(Q_b^i)$. As $\mathrm{cr}_D(C_4^{i-1}(G), G^i) = 0$ and $\mathrm{cr}_D(C_4^i(G), G^{i-1}) = 0$, the path $c_{i-1}b_{i-1}d_{i-1}$ cannot be crossed by an edge of G_b^i and the path $c_ib_id_i$ cannot be crossed by an edge of G_b^{i-1}. So, the unique possible crossing between G_b^{i-1} and G_b^i is a crossing between the edges $a_{i-1}b_{i-1}$ and a_ib_i. But, a crossing among the edges incident with the vertives a_{i-1}, a_i, and a_{i+1} is not the forced crossing in the subdivision Q_b^i of $K_{3,3}$. So, both forced crossings in $D(G_b^i)$ and $D(G_e^i)$ contribute together the value 2 to the force $f(Q^i)$, which implies that the total force of the drawing D is at least $2(n-1)$. This completes

the proof. ■

Theorem 1: $cr(G \square P_n) = 2(n-1)$ for $n \geq 1$.

Proof: Figure 2 shows the good drawing of the graph $G \square P_n$ with 2 crossings on every copy G^i, $i = 1, 2, \ldots, n-1$. Clearly, $cr(G \square P_n) \leq 2(n-1)$.

The inverse inequality is proved by induction with respect n. It is easy to see that $cr(G \square P_1) = 0$. By Lemma 1, $cr(G \square P_2) = 2$. So, this Theorem is true for $n = 1$ and $n = 2$.

The induction hypothesis is that theorem is true for $n = k, k \geq 2$. Suppose a good drawing D of $G \square P_{k+1}$ with fewer than $2((k+1)-1)$ crossings. According to Lemma 2, there exists a subgraph $G^i, i \in \{0, 1, \ldots, n\}$, with more than 1 crossing on its edges. The removing of all edges of this subgraph from D provides the drawing of a subdivision of $G \square P_k$ with fewer than $2(k-1)$ crossings or the drawing of a graph containing $G \square P_k$ as a subgraph, again with fewer than $2(k-1)$ crossings. This contradiction with the induction hypothesis completes the proof. ■

ACKNOWLEDGMENT

This work was supported by the Slovak VEGA grant No. 1/0309/11 (50%). This work is also the result of the project implementation Development of Center of Information and Communication Technologies for Knowledge Systems (project number: 26220120030) supported by the Research & Development Operational Program funded by the ERDF (50%).

REFERENCES

[1] C. D. Thompson, "Area-time complexity for VLSI," in *Proceedings of the eleventh annual ACM symposium on Theory of computing*, ser. STOC '79. New York, NY, USA: ACM, 1979, pp. 81–88.
[2] P. Turán, "A note of welcome," *J. Graph Theory*, vol. 1, pp. 7–9, 1977.
[3] K. Zarankiewicz, "On a problem of p.turán concerning graphs," *Fund. math.*, vol. 41, pp. 137–145, 1954.
[4] D. J. Kleitman, "The crossing number of $K_{5,n}$," *Journal of Combinatorial Theory*, vol. 9, no. 4, pp. 315–323, 1970.
[5] D. R. Woodall, "Cyclic-order graphs and Zarankiewicz's crossing-number conjecture," *J. Graph Theory*, vol. 17, no. 6, pp. 657–671, Dec. 1993.
[6] R. Guy, "Latest results on crossing numbers," in *Recent Trends in Graph Theory*, ser. Lecture Notes in Mathematics. Springer Berlin / Heidelberg, 1971, vol. 186, pp. 143–156, 10.1007/BFb0059432.
[7] S. Pan and R. B. Richter, "The crossing number of K_{11} is 100," *Journal of Graph Theory*, vol. 56, no. 2, pp. 128–134, 2007.
[8] D. Bokal, "On the crossing numbers of cartesian products with paths," *J. Combin. Theory Ser. B*, vol. 97, p. 2007, 2005.
[9] F. Harary, P. C. Kainen, and A. J. Schwenk, "Toroidal graphs with arbitrarily high crossing numbers," *Nanta Math.*, vol. 6, pp. 58–67, 1973.
[10] L. W. Beineke and R. D. Ringeisen, "On the crossing numbers of products of cycles and graphs of order four," *Journal of Graph Theory*, vol. 4, no. 2, pp. 145–155, 1980.
[11] S. Jendroľ and M. Ščerbová, "On the crossing numbers of $S_m \times P_n$ and $S_m \times C_n$," *Časopis pro pěstování matematiky*, vol. 107, pp. 225–230, 1982.
[12] M. Klešč, "The crossing numbers of products of paths and stars with 4-vertex graphs," *J. Graph Theory*, vol. 18, no. 6, pp. 605–614, Oct. 1994.
[13] M. Klešč, "The crossing numbers of cartesian products of paths with 5-vertex graphs," *Discrete Math.*, vol. 233, no. 1-3, pp. 353–359, Apr. 2001.
[14] M. Klešč, "The crossing number of $K_5 \times P_n$," *Tatra Mt. Math. Publ.*, vol. 18, pp. 63–68, 1999.

Title:

Electrical Engineering and Informatics III

Proceeding of the Faculty of Electrical Engineering and Informatics
of the Technical University of Košice

Editorial Board:
Chairman:
Dušan Kocur

Members:
Ján Kollár
Dušan Levický
Alena Pietriková
Ján Sarnovský
Liberios Vokorokos

Technical Editorial Board:
Marek Čajkovský
Katarína Kubišová
Jana Trelová

Published by:	**Faculty of Electrical Engineering and Informatics, Technical University of Košice, Letná 9, 042 00 Košice, Slovak Republic**
Year of publication:	**September, 2012**
Design:	**FEI TU of Košice**
Pages:	**836**
Edition:	**First edition**
Printing:	**50 pcs CD**
Language:	**English, Slovak**

Copyright © 2012 FEI TUKE

ISBN 978-80-553-0890-6