

Technical University of Košice



**Faculty of Electrical Engineering
and Informatics**

Electrical Engineering and Informatics

13

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

ISBN 978-80-553-4120-0

2022

Electrical Engineering and Informatics 13
Proceedings of the Faculty of Electrical Engineering and Informatics
of the Technical University of Košice

Published by: Faculty of Electrical Engineering and Informatics
Technical University of Košice
Letná 9, 040 01 Košice, Slovak Republic

Date of publication: August 2022 Language: English, Slovak
Printing: 50 pieces CD Pages: 561

Editorial board chairman: Alena Pietriková

Proceedings reviewers:

Norbert Ádám	Milan Lacko
František Babič	Branislav Madoš
Ján Bačík	Ján Magyar
Peter Bednár	Marián Mach
Peter Bober	Kristína Machová
Samuel Bucko	Dušan Medveď
Marek Bundzel	Miroslav Michalko
Zsolt Čonka	Jakub Oravec
Lubomír Doboš	Ján Papaj
Miloš Drutarovský	Ján Paralíč
František Ďurovský	Marek Pavlík
Jaroslav Džmura	Daniela Perduková
Peter Fecilak	Ján Perháč
Anna Filasová	Emília Pietriková
Ján Gamec	Matúš Pleva
Peter Girovský	Branislav Sobota
Iveta Gladišová	Ján Staš
Peter Gnip	William Steingartner
Daniel Hládek	Matúš Sulír
Ján Hurtuk	Csaba Szabó
Eva Chovancová	Martina Szabóová
Ondrej Kainz	Ján Šaliga
Erik Kajáti	Slavomír Šimoňák
Ján Kaňuch	Ján Vaščák
Jozef Király	Tibor Vince
Štefan Korečko	Marcel Vološin
Natalia Kurkina	Ján Zbojovský
Karol Kyslan	Iveta Zolotová
Viera Maslej Krešňáková	Jaroslava Žilková

Editors: Alena Pietriková
Emília Pietriková

Contents

Peter Girovský, Tomáš Matej <i>Návrh riadenia automatizovaného skleníka</i>	8
Peter Girovský, Patrik Pekala <i>Návrh ovládania 3D skenovacieho zariadenia</i>	13
Kristián Mičko, Peter Papcun <i>Použitie počítačového videnia na ovládanie IoT zariadení pre hendikepovaných ľudí</i>	17
Viktor Petro <i>Bezsnímačové riadenie synchrónneho motora s permanentnými magnetmi v oblasti nízkych rýchlostí</i>	21
Marek Pavlík, Dávid Dzúr <i>Sledovanie Parametrov Fotovoltaických Článkov</i>	25
Radoslav Figura, Branislav Madoš, Anton Baláž <i>Local currency creation using blockchain technology</i>	31
Michal Kucan, František Jakab, Dávid Cymbalak, Roman Vápeník <i>Optimalizácia topológií rozsiahlych sietí s efektívnou distribúciou komunikačných a streamingových služieb</i>	37
Gabriel Baran, Dušan Medved' <i>Open-source platformy pre elektroenergetické účely</i>	42
Ján Vaščák <i>Use of Recurrent Fuzzy Neural Networks in Control of Mobile Robots</i>	48
Branislav Sobota, Marián Hudák, Matej Kováč, Štefan Korečko, Miriama Mattová <i>Rozšírenie kolaboratívneho virtuálneho prostredia o podporu interaktívnej gestikulácie a vizualizácie rúk</i>	54
Maroš Lapčák, Ľuboš Ovseník, Jakub Oravec, Norbert Zdravecký, Samuel Andrejčík <i>Skúmanie dostupnosti a chybovosti FSO systému vplyvom hmly</i>	62
Branislav Sobota, Miriama Mattová, Lukáš Jasenka, Štefan Korečko <i>Modelovanie dronov v kolaboratívnom virtuálnom prostredí</i>	67
Ema Sujová, Martin Chovanec, Eva Chovancová <i>Visualisation of highway rest stops data via mobile application</i>	74
Martin Mattoš, Viera Anderková, František Babič <i>Analýza Výskytu Tsunami Pomocou Systému na Podporu Rozhodovania</i>	81
Eva Mročová, Branislav Madoš, Anton Baláž <i>Windows Service for preprocessing and static analysis of application logs of X-ray devices</i>	86
Jakub Oravec, Ľuboš Ovseník, Maroš Lapčák, Norbert Zdravecký, Samuel Andrejčík <i>Overenie vlastností logistického zobrazenia pomocou odhadu Ljapunovových exponentov</i>	92
Erik Andraško, Dušan Medved' <i>Sebestačnosť elektroenergetiky Slovenska</i>	97

Iveta Gladišová <i>Využitie Householderovej transformácie pri spracovaní hraničných blokov objektov</i>	102
Šimon Gans, Ján Molnár <i>A measurement method of basic ferromagnetic material properties</i>	107
Lucia Gojdičová, Michal Kolárik <i>Princípy syntetizácie lekárskeho správ</i>	113
Peter Girovský, Jaroslav Murín <i>Detekcia kolízny stavov motorizovaných zostáv</i>	118
Dávid Martinko, Michal Kolcun <i>Vplyv rôznych penetrácií domácich FVE na nízkonapäťovú distribučnú sieť</i>	123
Luboš Ovseník, Jakub Oravec, Norbert Zdravecký, Samuel Andrejčík, Maroš Lapčák <i>Skúmanie modulačných formátov pre Ultra-DWDM prenosové systémy</i>	128
Natalia Kurkina, Ján Papaj, Maroš Baumgartner <i>Problematika smerovania v Cloud MANET</i>	134
Ján Vaščák <i>Využitie evolučného ANFIS-u pre navigačné účely</i>	138
Martin Sedliak, Martina Szabóová <i>Detekcia sarkazmu na sociálnych siet'ach</i>	142
Miriama Jackaninová, Ján Hurtuk <i>Comparing the Performance of Current Tools for Dynamic Code Analysis</i>	146
Marek Pástor, Makar Golovkin <i>Využitie SPICE modelov v prostredí Matlab/Simulink</i>	153
Norbert Zdravecký, Luboš Ovseník, Maroš Lapčák, Jakub Oravec, Samuel Andrejčík <i>Návrh Ultra-DWDM systému s optickým zosilňovačom SOA</i>	159
Stanislav Ivanchyk, Martina Szabóová <i>Predicting Human Behaviour in Online Games Using Neural Networks</i>	165
Michal Michňák, Viera Anderková, František Babič <i>Analýza Nehodovosti na Cestách v USA Pomocou Vhodných Metód Dátovej Analytiky</i>	171
Tomáš Miščík, Marián Mach <i>Použitie algoritmu inšpirovaného svätajánskymi muškami pre numerickú optimalizáciu</i>	176
Natália Onufráková, Viera Anderková, František Babič <i>Podpora Diagnostiky Diabetes Mellitus Pomocou Vhodných Metód Dátovej Analytiky</i>	182
Adrián Marcinek, Marek Pástor <i>Návrh regulátorov a simulačné overenie výkonového multiport meniča</i>	187
Daniela Hudiková, Matúš Sulír <i>Comparison of Questions and Answers on Stack Exchange Sites of Content Management Systems</i>	194
Ivan Bank-Bedei, Marián Mach <i>Encoding Connect-4 using Quantified Boolean Formulae Revisited</i>	200

Zuzana Pugelová, Ján Jadlovský <i>Design and Implementation of a Web Application for the CMCT&II Research Group</i>	206
Vojtech Kalazi, Branislav Madoš, Anton Baláž <i>Voxelizačný algoritmus generujúci SVDAg hierarchickú dátovú štruktúru</i>	212
Maroš Baumgartner, Natalia Kurkina, Ján Papaj <i>Robustný Prenos Dát v 5G Sieťach bez Infraštruktúry Založený na Technológii Blockchain</i>	218
Jakub Oravec, Ľuboš Ovseník, Maroš Lapčák, Norbert Zdravecký, Samuel Andrejčík <i>Použitie štatistických testov na overenie vlastností obrazových šifrovacích algoritmov</i>	224
Jozef Bučko, Matúš Pleva <i>Analýza a klasifikácia behaviorálnych biometrických dát</i>	228
Branislav Sobota, Miriama Mattová, Róbert Fedorco, Štefan Korečko <i>Training and testing system for a humanoid robot in a virtual reality</i>	234
Martin Murcko, František Jakab, Dávid Cymbalak, Roman Vápeník <i>Data flow analysis of videoconference devices in network infrastructure</i>	241
Stanislav Ondáš, Michaela Novotná <i>Prehliadač n-gramových jazykových modelov</i>	249
Patrik Zelenák, Miloš Drutarovský <i>DLL knižnica hašovacej funkcie Blake3 pre Python</i>	254
Ján Kaňuch, Ivana Savinecká <i>Návrh a riadenie inteligentného skleníka</i>	260
Marek Pástor, Jaroslav Dudrik, Richard Michal <i>Nepriamy jednosmerný menič s jednoduchým odľahčovacím obvodom na sekundárnej strane</i>	270
Dmytro Lahunov, Ivan Čík, Marián Mach <i>Necielové kontradiktórne útoky na klasifikačné hlboké modely</i>	276
Dávid Lacko, Marián Mach <i>Prvé experimenty s hracím agentom založeným na algoritme MCTS</i>	282
Daniel Majirský, Anton Baláž, Branislav Madoš <i>Skúmanie malvéru a neškodného softvéru v dokumentoch PDF</i>	287
Ján Galovič, Martin Hasin, Martin Havrilla <i>Tool for generating UI schemes</i>	295
Simona Kirešová, Milan Guzan <i>Koncentrácia pevných častíc v ovzduší: porovnanie meraní</i>	301
Samuel Andrejčík, Ľuboš Ovseník, Jakub Oravec, Norbert Zdravecký, Maroš Lapčák <i>Digitálna vodotlač zabezpečujúca obrazové dáta pred útokmi vyrezaním</i>	307
Jakub Frankovič, Ondrej Kainz, Miroslav Michalko, František Jakab, Ivana Nováková <i>Web-based Solution for Management of Records</i>	311
Ján Vaščák <i>Neurónové siete typu RBF pri navigácii robota</i>	319
Norbert Zdravecký, Ľuboš Ovseník, Samuel Andrejčík, Jakub Oravec, Maroš Lapčák <i>Návrh 0,32 Tb/s Ultra-DWDM systém s optickým zesilňovačom ROA</i>	325

Maroš Baumgartner, Ján Papaj, Erik Sarsala <i>Model D2D Komunikácie pre 5G Siete</i>	330
Bianka Bilecová, Ján Hurtuk <i>Online Voting Management System Based on Blockchain Technology</i>	335
Marek Glevaňák, Martin Hasin, Martin Chovanec, Martin Havrilla <i>Design optimized flow collector for network capturing</i>	340
Stanislav Ondáš, Silvia Bačíková <i>Automatická detekcia miest relevantných pre zmenu rečníka v dialógových interakciách</i>	347
Patrik Schwartz, Ondrej Kainz, Miroslav Michalko, František Jakab, Ivana Nováková <i>Portal for automated quality assessment of goods</i>	352
Marek Pavlík, Dávid Olšinár <i>Porovnanie Návrhov Fotovoltaiky na Rodinný Dom</i>	362
Martin Binar, Dušan Medved <i>Návrh netypizovaného energetického zdroja a jeho praktické využitie</i>	369
Branislav Vangor, Samuel Bucko <i>Ekonomické a environmentálne aspekty decentralizovanej výroby elektriny</i>	375
Júlia Dulinová, Anton Baláž, Emília Pietriková <i>Riešenie bezpečnosti informačných technológií pomocou hier</i>	381
Kristína Machová, Martin Kaňuch <i>Analýza regresných metód z hľadiska vhodnosti na detekciu trollov</i>	387
Marek Roháč, Miloš Drutarovský <i>Kryptografický balík Xoodyak a jeho použitie v jednoduchej VPN sieti</i>	391
Tomáš Horčík, Ján Papaj <i>Softvérovo-definované MANET siete</i>	397
Dominik Tkáčik, Csaba Szabó, Róbert Vandlík <i>Detekcia kvality orezu moriacich liniek</i>	403
Samuel Andrejčík, Ľuboš Ovseník, Jakub Oravec, Maroš Lapčák, Norbert Zdravecký <i>Návrh mobilnej aplikácie pre získavanie dát z monitorovacej stanice</i>	409
Tomáš Hudák, Marián Mach <i>Evolúcia obrazu použitím evolučného algoritmu</i>	415
Kristian Lutak, Slavomír Šimoňák <i>Supporting the Education of Algorithms and Data Structures Through Educational Games</i>	421
Stanislav Ondáš, Lea Šimčíková <i>Technologická podpora logopedickej terapie v slovenčine</i>	427
Kristína Machová, Michal Porezaný <i>Algoritmy strojového učenia v identifikácii trolizmu</i>	433
Tímea Balhová, Richard Lonščák, Anna Jadlovska <i>Predikčné finančné modely časových radov s využitím vlnkových transformácií</i>	437
Adam Račko, Jana Fortes <i>Softvérová realizácia zobrazovacej jednotky UWB senzorovej siete</i>	443

Anita Tóthová, Anton Baláž, Branislav Madoš <i>Computer Security through Gamification</i>	449
Dominika Culbová, Viera Anderková, František Babič <i>Pochopenie Vzniku Tsunami Pomocou Metód Dátovej Analytiky</i>	457
Ján Vaščák <i>Robot Control in the Environment of Virtual Reality</i>	463
Marek Fedor, Daniela Perduková <i>Univerzálny emulátor energetických systémov</i>	468
Gabriela Vašková, Natalia Kurkina, Ján Papaj, Maroš Baumgartner <i>Alternatívne metódy smerovania pre 6G siete</i>	476
Marek Pástor, Jaroslava Žilková <i>Napáťový regulátor pre nepriamy jednosmerný menič s mäkkým spínaním</i>	480
Luboš Ovseník, Jakub Oravec, Norbert Zdravecký, Maroš Lapčák, Samuel Andrejčík <i>Návrh a realizácia FTTH siete</i>	485
Lubomír Rusnák, Marián Mach <i>SAT kódovanie kardinalitných ohraničení pomocou P-H metódy</i>	491
Maroš Baumgartner, Natalia Kurkina, Ján Papaj, Richard Neznik <i>Secure routing in 6G network</i>	496
Ján Vaščák <i>Visual Interaction with Intelligent Space</i>	501
Peter Girovský, Tadeáš Kmecik <i>Lineárny pohon</i>	507
Peter Girovský, Viktor Gamrát <i>Návrh GUI pre ovládanie pohonu</i>	512
Matej Kormanský, Branislav Madoš <i>Analýza vplyvu zmeny dimenzionality na efektívnosť reprezentácie geometrie voxelizovaných 3D scén</i>	517
Maroš Lapčák, Luboš Ovseník, Jakub Oravec, Norbert Zdravecký, Samuel Andrejčík <i>Návrh mikropásikovej antény pre použitie v hybridnom FSO/RF systéme</i>	523
Kristína Machová, Natalia Dankanych <i>Identifikácia rôznych typov správ o CoViD-19 pomocou zhlukovacích metód</i>	528
Vladimír Jacko, Branislav Madoš, Anton Baláž <i>Využitie algoritmu počítačového videnia v priemyselnej aplikácii</i>	532
Ján Papaj, Erik Seidenglanz <i>Technológia blockchain a smerovací protokol AODV v MANET</i>	539
Kristína Machová, Michaela Horváthová <i>Výber relevantných atribútov asociálneho chovania</i>	543
Kristína Machová, Samuel Široký <i>Identifikácia trollov na základe podobnosti textov</i>	547

Pavol Fedor, Peter Radváni <i>An Inverse Fuzzy Model Strategy for Dynamic System Control</i>	552
Author's Index	559

Návrh riadenia automatizovaného skleníka

¹Peter GIROVSKÝ, ²Tomáš MATEJ

^{1,2} Katedra elektrotechniky a mechatroniky, Fakulta elektrotechniky a informatiky,
Technická univerzita v Košiciach

¹peter.girovsky@tuke.sk, ²tomas.matej@student.tuke.sk

Abstrakt — Cieľom tejto práce je predstaviť návrh riadenia automatizovaného skleníka, ktorý bude schopný udržiavať stabilné podmienky vhodné pre pestovanie rastlín. Návrh riadenia pozostával z troch častí, analýza riešenia automatizovaného skleníka, návrh komponentov a samotný návrh algoritmu riadenia automatizovaného skleníka.

Kľúčové slová — skleník, riadenie, automatizácia, Arduino, snímače

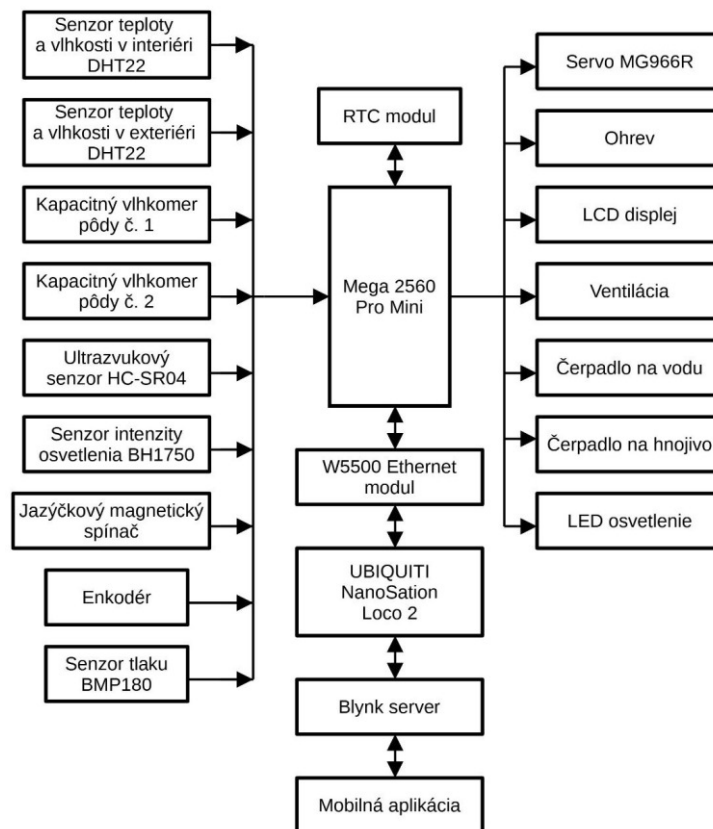
I. ÚVOD

Riešenia v oblasti automatizovaného riadenia skleníka sa zameriavajú hlavne na udržiavanie optimálnej teploty vzduchu a vlhkosti pôdy v skleníku, keďže sú to najviac ovplyvňujúce faktory. Každé riešenie však prináša aj niečo navyše k týmto základným funkciám, pričom pre riadenie sa používajú rôzne riadiace systémy a rôzne snímače s rozdielnymi vlastnosťami.

V článku [1] autori navrhli jednoduchý, ľahko inštalovateľný systém, ktorý sa skladá zo senzorov snímajúcich vlhkosť pôdy, teplotu a vlhkosť vzduchu, intenzitu osvetlenia. Namerané údaje načítaval mikrokontrolér Atmega328, ktorý riadi skleník. Toto riešenie udržiava optimálne prostredie vhodné pre efektívne pestovanie plodín v skleníku. Mikrokontrolér na základe parametrov nameraných snímačmi a internej logiky, ktorá je do neho vpísaná, vyhodnotí aktuálnu situáciu v skleníku. Namerané údaje neustále porovnáva s prahovými hodnotami zadefinovanými v programe a na základe týchto porovnaní rozhoduje, či je potrebné aktivovať akčné členy, ktoré by upravili merané veličiny v skleníku. Autori článku [2] využili na svoju prácu dosku Netduino 3 Wi-Fi. Tá poskytuje výkonnejší hardvér vďaka procesoru Cortex-M4, viac flash a RAM pamäti. Má zabudované pripojenie Wi-Fi a Bluetooth. Takto navrhnutý systém automatizovaného skleníka založený na IoT umožňuje spojenie medzi ľuďmi a vecami, ktoré ich obklopujú. V tomto prípade umožňuje jednoduchšie zhromažďovanie informácií zo skleníka v reálnom čase, ich analýzu, spracovanie a následne samotné riadenie skleníka. Všetky potrebné informácie pre automatizované riadenie tohto skleníka sa prenášali prostredníctvom Google cloud IoT. Takto sú informácie o skleníku a jeho riadenie prístupné odkiaľkoľvek a kedykoľvek, čo značne uľahčí prácu pestovateľom. V článku [3] je systém automatizovaného skleníka riadený doskou Arduino UNO. To je prepojené so senzormi: osem kanálovým relé modulom, časovým modulom a Wi-Fi modulom. Celý systém je rozdelený na riadiacu a monitorovaciu časť. V práci [4] sa zamerali na kontrolovanie klimatických podmienok pre optimálnu produkciu plodín v skleníku. Celý systém sa skladal z dosky Arduino Mega 2650, snímačov teploty a vlhkosti vzduchu, vlhkosti pôdy a intenzity osvetlenia. Využívali aj senzory prietoku kvapaliny, plynové senzory, prúdový snímač a aktuátory ako relé, čerpadlo a elektromagnetický ventil. Skleník, v ktorom systém implementovali a testovali mal rozlohu 16 m². Jedna z požiadaviek na tento monitorovací systém bola, aby jednotkové náklady na tento monitorovací systém boli nízke.

II. NÁVRH MODELU AUTOMATIZOVANÉHO SKLENÍKA

Návrh modelu automatizovaného skleníka bol robený z hľadiska možnosti jeho implementácie aj do väčšieho skleníka s použitím výkonnejších akčných členov a zároveň, aby takéto riešenie ponúkalo dobrý pomer ceny a kvality. Blokovaná schéma navrhnutého systému je na Obr. 1.



Obr. 1 Bloková schéma navrhnutého systému

A. Zavlažovanie substrátu

Pre optimálny rast rastlín je potrebné udržiavať substrát dostatočne vlhký. Spúšťanie zavlažovacieho čerpadla je v tomto návrhu závislé od vlhkosti substrátu, v ktorom budú umiestnené rastliny. Vlhkosť pôdy bola meraná kapacitným vlhkomerom pri ktorom nedochádza ku korózii kovových kontaktov umiestnených v pôde, ako je tomu v prípade snímača YL69. V návrhu boli použité dva snímače vlhkosti, pričom každý z týchto senzorov prevedie nameranú hodnotu na analógový signál. Doska Arduino Mega 2560 PRO tieto hodnoty spriemeruje a na základe neho vyhodnotí, či je potrebná zálievka substrátu, alebo nie.

Závlaha bude možná iba v prípade, ak bude dostatok vody v nádrži. Na zisťovanie stavu nádrže sú na trhu k dispozícii rôzne plavákové senzory, no tie kvôli svojim vlastnostiam nie sú vyhovujúce pretože neumožňujú merať spojito stav naplnenia nádrže. Preto bol pre snímanie vody v nádrži zvolený ultrazvukový snímač vzdialenosti HC-SR04.

B. Regulácia teploty v skleníku

Stabilnú teplotu vhodnú pre rastliny umiestnené vo vnútri skleníka zabezpečí jeho ventilácia. V návrhu je na meranie teploty použitý senzor DHT22. Nameraná hodnota vo forme digitálneho signálu je zaslaná Arduino doske, ktorá ju spracuje a porovná s maximálnou a minimálnou dovolenou teplotou. Na základe tohto porovnania sa vykoná akcia zabezpečujúca úpravu teploty na požadovanú hodnotu. Vetranie v skleníku bude rozdelené do dvoch krokov, v prvom kroku vetrania bude otvorené okno pomocou servo motora MG966R. Pokiaľ sa nepodarí týmto spôsobom teplotu znížiť, ale naopak teplota s časom bude naďalej stále rásť, tak sa spustí elektrická ventilácia spínaná relé modulom, ktorá pozostáva z dvoch axiálnych ventilátorov s priemerom 60mm napájaných na 12 voltov. Tá bude v prevádzke pokiaľ sa teplota nedostane pod požadovanú hodnotu.

C. Regulácia osvetlenia

Umelé osvetlenie bolo realizované na základe požiadavky dostatočnej intenzity svetla pre optimálny rast rastlín, pričom intenzita svetla je meraná pomocou senzora intenzity svetla BH1750.

D. Grafický LCD displej

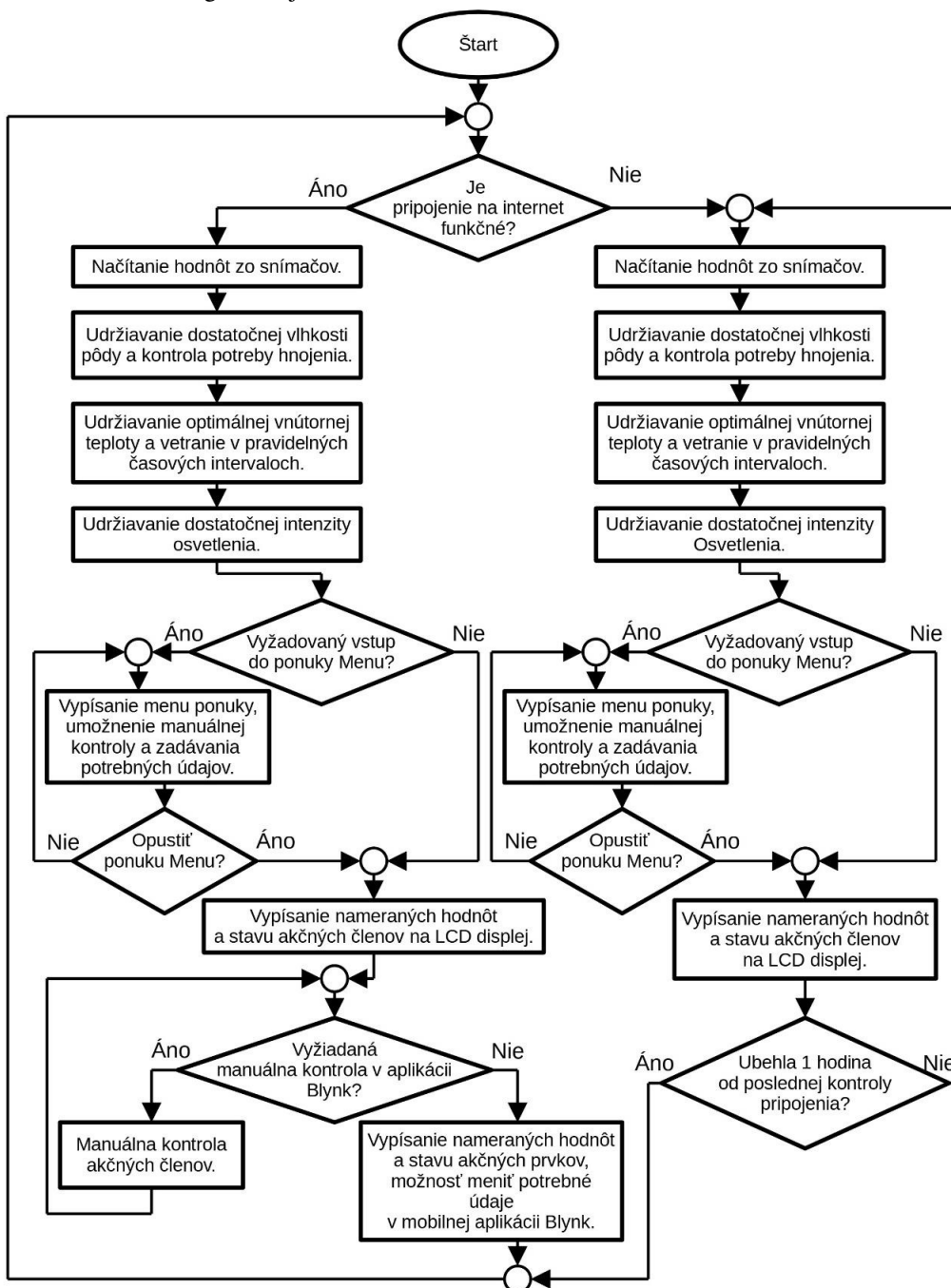
Bude slúžiť na informovanie používateľa o aktuálnom stave meraných veličín a taktiež akčných členov.

E. Mobilná aplikácia

Pri výbere mobilnej aplikácie bolo hlavným kritériom schopnosť zobrazovať namerané údaje zo skleníka, riadiť jeho akčné členy a zároveň, aby všetky tieto funkcie boli dostupné odkiaľkoľvek prostredníctvom internetu. Na základe oho bola zvolená aplikácia Blynk, čo je IoT platforma, ktorá umožňuje komunikáciu zariadení prostredníctvom internetu, ich riadenie a zhromažďovanie nameraných dát. Mobilnú aplikáciu si je možné prispôsobiť vlastným potrebám pridávaním a úpravou blokov s rôznymi funkciami. Pre jej správne fungovanie je potrebná mobilná aplikácia, Blynk server a hardvér schopný komunikovať s týmto serverom pomocou vytvorených knižníc.

III. NÁVRH RIADIACEHO ALGORITMU

Úlohou riadiaceho algoritmu je udržať optimálne podmienky na pestovanie a informovať o aktuálnom stave meraných veličín a akčných členov. Navrhnutý algoritmus bude udržiavať tri základné faktory - teplotu, vlhkosť pôdy a dostatočnú intenzitu svetelného žiarenia. Bloková schéma riadiaceho algoritmu je na obr. 2.



Obr. 2 Bloková schéma navrhnutého riadiaceho algoritmu

IV. POROVNANIE VLASTNOSTÍ NAVRHNUTÉHO SKLENÍKA S DOSTUPNÝMI RIEŠENIAMÍ

Po skompletovaní fyzikálneho modelu skleníka zobrazeného na obr. 3 bola overená jeho plná funkčnosť vrátane riadenia a všetkých jeho podsystémov.



Obr. 3 Model skleníka

Porovnanie vlastností návrhu automatizovaného skleníka s dostupnými riešeniami uvedenými v úvode článku, je zobrazené v tabuľke 1.

Tabuľka 1

Porovnanie vlastností navrhnutého riešenia s existujúcimi riešeniami uvedenými v úvode práce

	Automatická závlaha	Automatické vetranie	Umelé osvetlenie	Užívateľské rozhranie	Kontrola na diaľku	Doplnkové funkcie
Riešenie 1	Na základe odmeranej vlhkosti pôdy.	Založené na základe odmeranej teploty a vlhkosti.	Spínané v dvoch stavoch, bez regulácie.	Dvojriadkový LCD displej s informačným charakterom.	Nie je k dispozícii.	Bez doplnkových funkcií.
Riešenie 2	Na základe odmeranej vlhkosti pôdy senzorom YL69.	Založené na základe odmeranej teploty.	Založené na základe odmeranej intenzity.	Dvojriadkový LCD displej a IOT platforma.	Umožňuje diaľkovú kontrolu a riadenie.	SMS notifikácie.
Riešenie 3	Fixne daná v čase, nezávislá od vlhkosti pôdy.	Viacstupňové, založené na odmeranej teplote.	Bez regulácie spínané v dvoch stavoch.	Mobilná aplikácia Virtuino.	Neumožňuje riadenie iba kontrolu stavu.	Email notifikácie.
Riešenie 4	Je súčasťou riešenia, no z práce nie je zrejme od čoho je závislá.	Okrem ventilácie obsahuje aj vykurovanie.	Bez umelého svietenia.	16 znakový LCD displej s informačným charakterom a webová stránka.	Webová stránka umožňuje priame riadenie a kontrolu stavu skleníka.	Snímanie kvality ovzdušia.
Vlastné riešenie	Na základe odmeranej vlhkosti pôdy.	Viacstupňové, založené na odmeranej teplote s pridaným kúrením.	S plynulou reguláciou, ktorá dorovnáva žiadanú intenzitu.	Grafický LCD displej a mobilná aplikácia, s možnosťou zmeny hraničných hodnôt.	Mobilná aplikácia umožňuje priame riadenie a kontrolu stavu skleníka.	Emailové a mobilné notifikácie o poruchách, meranie exteriérovej teploty a tlaku.

V. ZÁVER

Tento článok prezentoval návrh riadenia automatizovaného skleníka, ktorý by bol schopný

samostatne udržiavať optimálne podmienky pre pestovanie v skleníku a zároveň by podával informácie o aktuálnom stave prostredníctvom mobilnej aplikácie a LCD displeja. Na základe testovania navrhnutého konceptu skleníka môžeme konštatovať, že navrhnutý systém sa preukázal ako funkčný a spoľahlivý.

POĎAKOVANIE

Táto práca bola podporovaná Agentúrou na podporu výskumu a vývoja na základe Zmluvy č. APVV-19-0210.

Táto práca vznikla s podporou grantu VEGA 1/0493/19.

LITERATÚRA

- [1] SAHA, T. et al. Construction and Development of an Automated Greenhouse System Using Arduino Uno. In: International Journal of Information Engineering and Electronic Business [online]. 2017, roč. 9, č. 3, s. 1-8. ISSN 20749023. DOI: 10.5815/ijieeb.2017.03.01
- [2] S. RAJ, Jennifer, J, Vijitha Ananthi. Automation Using Iot in Greenhouse Environment. In: Journal of Information Technology and Digital World [online]. 2019, roč. 01, č. 01, s. 38-47. DOI: 10.36548/jitdw.2019.1.005
- [3] ROHIMI, R. et al. Design and prototype development of automated greenhouse with arduino and (IoT) application. In: International Journal of Advanced Science and Technology [online]. 2019, roč. 28, č. 16, s. 437-446. ISSN 22076360. Dostupné na internete: <https://www.researchgate.net/publication/338401070_Design_and_Prototype_Development_of_Automated_Greenhouse_with_Arduino_and_IoT_Application>
- [4] VOICU, Gheorghe, CONSTANTIN, Gabriel Alexandru. an Inexpensive Open-Source System for Monitoring and Control Environmental Factors in Greenhouse / Sistem Open-Source Cu Costuri Reduse Pentru Monitorizarea Și Controlul Factorilor ... an Inexpensive Open-Source System for Monitoring and Control. . 2015, č. May 2017
- [5] Kapacitný senzor vlhkosti pôdy [online]. [cit. 2021-5-18]. Dostupné na internete: <<https://bmes.com/product/capacitive-soil-moisture-sensor/>>
- [6] Ultrazvukový merač vzdialenosti HC-SR04 [online]. [cit. 2021-5-18]. Dostupné na internete: <<https://geeksville.com/en/product/ultrasonic-sensor/>>
- [7] Senzor teploty DHT22 [online]. [cit. 2021-5-18]. Dostupné na internete: <<https://arduino.co.ke/product/high-precision-am2302-dht22-digital-temperature-humidity-sensor-module-for-arduino-uno-r3/>>
- [8] Servo MG996r [online]. [cit. 2021-5-18]. Dostupné na internete: <<https://www.gme.sk/servo-mg996r-11kg-cm>>
- [9] Peltierov článok TEC1-12706 [online]. [cit. 2021-5-18]. Dostupné na internete: <<http://espressomilkcooler.com/wp-content/uploads/2015/03/TEC1-12706-site-ready.pdf>>
- [10] Peltierov článok TEC1-12706 [online]. [cit. 2021-5-18]. Dostupné na internete: <<https://diolut.sk/modul-peltier-tec1-12706-12v-60w-p-8747.html>>
- [11] Senzor intenzity svetla BH1750 [online]. [cit. 2021-5-18]. Dostupné na internete: <<https://bmes.com/product/bh1750-light-intensity-sensor/>>
- [12] Grafický LCD displej 128x64 ST7920 [online]. [cit. 2021-5-18]. Dostupné na internete: <https://diolut.sk/lcd-displej-12864-128x64-modr-5v-st7920-p-17784.html>
- [13] Ethernet modul W5500 [online]. [cit. 2021-5-18]. Dostupné na internete: <<https://leoricksimon.blogspot.com/2018/06/w5500-ethernet-module-with-arduino-mega.html>>
- [14] RTC modul DS3231 s pamäťovým modulom AT24C32 [online]. [cit. 2021-5-18]. Dostupné na internete: <<https://www.gme.sk/modul-rtc-ds3231-2c-a-32kb-flash>>
- [15] Modul BMP180 merajúci atmosférický tlak [online]. [cit. 2021-5-18]. Dostupné na internete: <<https://www.gme.sk/modul-meteocidlo-bmp180>>
- [16] Relé modul [online]. [cit. 2021-5-18]. Dostupné na internete: <<https://www.gme.sk/rele-modul-4-kanaly-10a-250v-opticke-oddelenie>>
- [17] Nastaviteľný regulátor znižujúci napätie [online]. [cit. 2021-5-18]. Dostupné na internete: <<https://www.gme.sk/napajaci-modul-3-40v-3a-step-down>>
- [18] UBIQUITI NanoStation Loco 2 [online]. [cit. 2021-5-18]. Dostupné na internete: <https://dl.ubnt.com/loco2_datasheet.pdf>
- [19] UBIQUITI NanoStation Loco 2 [online]. [cit. 2021-5-18]. Dostupné na internete: <https://datacomp.sk/ubiquiti-nanostation-loco-2-outdoor-2-4ghz-klient-vc-anteny_d50937.html>
- [20] Vývojová doska Mega 2560 PRO MINI [online]. [cit. 2021-5-18]. Dostupné na internete: <<https://www.makerlab-electronics.com/product/arduino-mega-2560-pro-mini-atmega2560-16au-ch340g/>>
- [21] MAUGHAN, Tiffany, DROST, Dan, BLACK, Brent. Supplemental Heating in a High Tunnel. In: [online]. 2014, č. July, s. 3-5. Dostupné na internete: <https://digitalcommons.usu.edu/cgi/viewcontent.cgi?referer=http://blogs.cornell.edu/hightunnels/structures/&https_redir=1&article=1678&context=extension_curall>
- [22] BALEMI, T et al. Cultivation of tomato: production, processing and marketing [online]. 2005. ISBN 9783540345350. DOI: 10.4067/S0718-95162012005000015
- [23] Mini skleníka využitý na realizáciu modelu [online]. [cit. 2021-5-18]. Dostupné na internete: <<https://www.ikea.com/sk/sk/p/socker-mini-sklenik-biela-70186603/>>
- [24] PARADISO, Roberta, PROIETTI, Simona. Light-Quality Manipulation to Control Plant Growth and Photomorphogenesis in Greenhouse Horticulture: The State of the Art and the Opportunities of Modern LED Systems [online]. Springer US, 2021. ISBN 0123456789. DOI: 10.1007/s00344-021-10337-y

Návrh ovládania 3D skenovacieho zariadenia

¹Peter GIROVSKÝ, ²Patrik PEKALA

^{1,2} Katedra elektrotechniky a mechatroniky, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach

¹peter.girovsky@tuke.sk, ²patrik.pekala@student.tuke.sk

Abstrakt — Cieľom tejto práce je predstaviť návrh riadenia ochranného krytu priemyselného zariadenia pre 3D skenovanie. Skenovacie zariadenie je potrebné ochrániť pred vplyvom vysokých teplôt z hľadiska jeho využívania. Samotné skenovacie zariadenie je tvorené dvoma funkčnými celkami a to samotným skenovacím zariadením a ochranným krytom. V rámci práce je predstavený návrh hardvérového a softvérového riešenie ochranného krytu, ktorý bol následne aj realizovaný.

Kľúčové slová — PLC, modbus, sériová linka, servomotor, vizualizácia, 3D skenovanie, 3D skener

I. ÚVOD

Mnohé procesy v ťažkom železiarskom priemysle obsahujú časti, pri ktorých existuje priestor na optimalizáciu, ktorou sa zabezpečí zvýšenie efektivity procesu pri zachovanej kvalite produktu, čo priamo vedie k zníženiu nákladov. Jedným z týchto procesov, pri ktorých je priestor na zlepšenie, je aj rovnomerné posypávanie horúcej roztavenej ocele práškom. Tento proces prebieha v desiatkach železiarní po celom svete a žiadna z nich ho nemá dostatočne zautomatizovaný. Problémom pri sypaní prášku je zabezpečenie rovnomernosti nanášanej vrstvy. Tejto úlohy sa často zhostuje operátor výroby, ktorý je tak vystavovaný vysokej teplote a veľmi nehostinnému prostrediu. Zároveň však nemá dostatočne dobrý dosah a možnosti, ako dokonale vyhladiť vrstvu prášku. Keby existovala možnosť, ako spoľahlivo naskenovať celú plochu tekutej ocele zasypávanej práškom, identifikácia nedokonalostí vrchnej vrstvy by bola jednoduchšia. Toto meranie by eventuálne mohlo tvoriť základný kameň automatického systému sypania prášku, ktorý by viac nevystavoval operátora výroby ťažkým pracovným podmienkam a zároveň účinne šetril materiál.

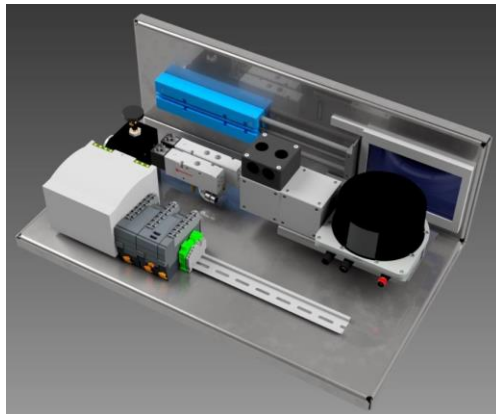
Na základe reálnej požiadavky z výrobného odvetvia bol vytvorený základný koncept snímacieho zariadenia, ktoré je schopné vytvoriť presný 3D model snímanej plochy a zvýrazniť miesta, ktoré si vyžadujú väčšie množstvo materiálu. Tento koncept bol rozdelený do dvoch samostatných celkov ktoré bolo potrebné riešiť. Prvý celok je zameraný na ovládanie, konfiguráciu a zber údajov z 3D skenera a druhý je zameraný na ukladanie údajov zo skenera, ich spracovanie, preškáľovanie na vhodný rozsah, zobrazovanie údajov na 3D obrázku a ovládanie pohybu skenera, i samotný manažment stavu meracieho zariadenia, pričom realizácia tejto druhej časti je predstavená v tomto príspevku.

II. NÁVRH HARDVÉROVEJ KONFIGURÁCIE

Hardvér skenovacieho zariadenia, zobrazený na obr. 1, pozostáva z dvoch samostatných častí, a to z ochranného krytu skenovacieho zariadenia a z samotného skenovacieho zariadenia. Skenovacia časť pozostáva z 3D skenera, ktorého signály sa spracovávajú pomocou Raspberry Pi a z časti ochranného krytu, ktorá zabezpečuje ochranu 3D skenera pred nepriaznivými prevádzkovými podmienkami a samotné polohovanie 3D skenera, pričom táto časť zariadenia je riadená pomocou PLC. Komunikačným rozhraním medzi riadiacimi systémami, teda medzi Raspberry Pi a PLC, je Modbus server, ktorý je prevádzkovaný v PLC a Raspberry pristupuje na jeho registre. PLC odosiela a prijíma signály zo servomotora využitím klasickej jednoduchej trojvodičovej sériovej linky.

Nosnou časťou hardvérovej konfigurácie pre ovládanie ochranného krytu 3D skenovacieho zariadenia sú:

- Servopohon SCHUNK 0307300 PR 070-161-B
- PLC B&R X20 CP 0482



Obr. 1 Hardvérová konfigurácia zariadenia

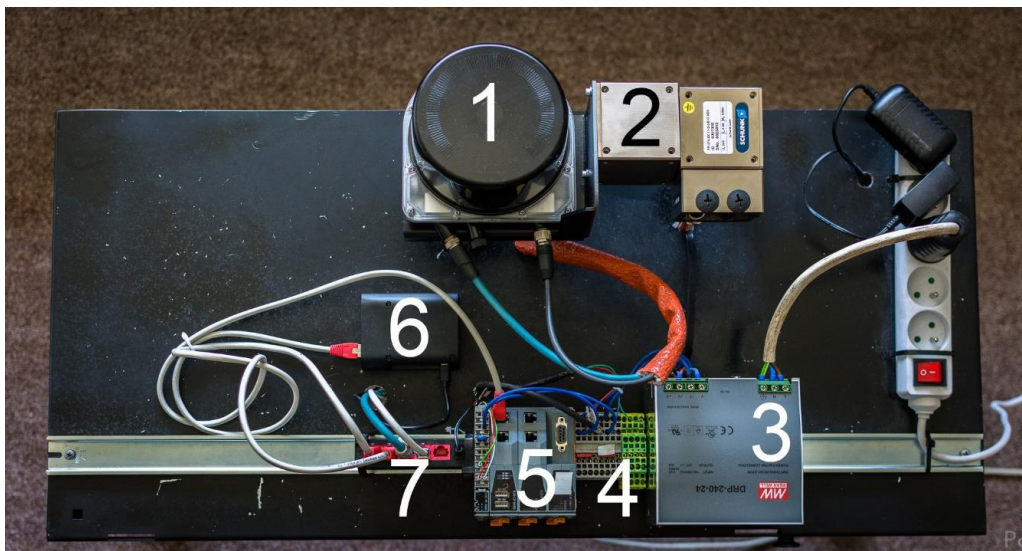
A. Servopohon SCHUNK 0307300 PR 070-161-B

Servopohon od firmy SCHUNK slúži na presné polohovanie 3D skenera pre snímání obrazu [2, 5].

Výkonová aj riadiaca časť tohto servopohonu je napájaná jednosmerným napätím 24V a servomotor disponuje magnetickou brzdou, ktorá je aktívna pri chybovom stave, pri výpadku napájania, či po každom dosiahnutí želanej polohy. Najdôležitejším parametrom riadenia presného pohonu je jeho opakovateľná presnosť polohovania, ktorá pri aktuálne použítom PLC dosahuje hodnotu na úrovni $0,1^\circ$.

B. PLC B&R X20 CP 0482

Pri voľbe vhodného riadiaceho zariadenia bol zvolený malý kompaktný PLC od firmy B&R, ktorého výpočtový výkon postačoval pre uvažované riadenie základných funkcií ochranného krytu 3D skenera [1], medzi ktoré patrí riadenie presného servopohonu SCHUNK s dostatočnou presnosťou polohovania, riadenie vizualizácie a komunikácie s Raspberry PI, ktoré zabezpečovalo spracovanie signálov z 3D skenera. Výkon uvedeného PLC je taktiež postačujúci, pre doplnenie ďalších funkcií ako je napr. riadenie priezoru ochranného krytu a riadenie teploty.



Obr. 2 Testovacia verzia 3D skenovacieho zariadenia

Na obr. 2 je možné vidieť testovaciu verziu skenovacieho zariadenia.

- 1 - 3D LiDAR skener SICK,
- 2 - servopohon SCHUNK 0307300 PR 070-161-B,
- 3 - napájací zdroj MEAN WELL DRP-240-24,

- 4 - svorky pre vyvedenie + a – zdroja a zemnenie,
- 5 - PLC,
- 6 - Raspberry Pi 3,
- 7 - sieťový switch.

III. NÁVRH A REALIZÁCIA SOFTVÉROVÉHO RIEŠENIA

Návrh programu pre danú úlohu bol robený v prostredí Automation Studio od firmy B&R. Riešenie niekoľkých čiastkových úloh a zadaní bolo rozdelené do programov, z ktorých každý nesie špecifickú časť celkovej implementácie riadenia. Projekt obsahuje nasledujúce časti:

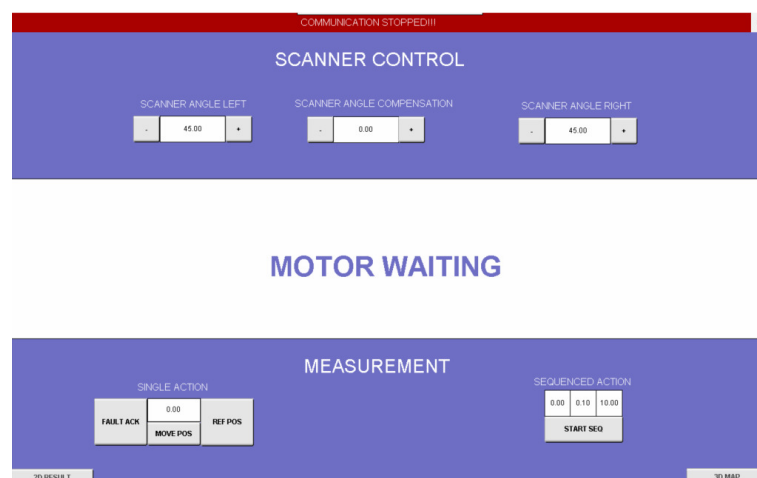
- komunikáciu so scannerom pomocou Modbus servera realizovanom na PLC
- sledovanie komunikácie so scannerom – watchdog
- komunikáciu s motorom prostredníctvom sériovej linky
- vizualizáciu
- program, zabezpečujúci správnu interpretáciu nameraných výsledkov.

IV. OVLÁDACÍ PANEL

Vizualizácia, ktorá umožňuje operátorovi zariadenie priamo ovládať, bola realizovaná pomocou VNC servera. Takéto riešenie umožňuje použiť na zobrazovanie širokú škálu zariadení, ktoré obsahujú potrebné vstupné a výstupné komponenty. Zariadenie potrebuje pripojenie do rovnakej siete ako PLC, prípadne priame spojenie ethernetom s PLC. Takýmto zariadením môže byť PC, smartfón, či tablet.

Pomocou realizovanej vizualizácie užívateľ nielenže má možnosť ovládať príslušné zariadenie obr. 3, ale vidí aj v akom stave sa nachádza motor a v rámci navrhutej vizualizácie sa zobrazujú aj namerané výsledky obr. 4. Ovládacia časť GUI obsahuje niekoľko prvkov, ktorými dokáže užívateľ riadiť proces merania. Okno je rozdelené do troch funkčných oblastí. V hornej tretine okna dokáže užívateľ riadiť parametre, ktoré PLC odosiela priamo scanneru. Premenné SCANNER ANGLE LEFT a SCANNER ANGLE RIGHT riadia „šírku“ merania. Predvolene meria snímač scénu širokú 90°- 45° vľavo od nulovej polohy a 45° vpravo od nulovej polohy. Predvolená hodnota je zároveň maximálnou hodnotou šírky meranej scény - užívateľ dokáže zvoliť iba uhol v rozsahu 0 – 45 ° vľavo, i vpravo. Stred ovládacieho okna zobrazuje stav motora. Informácia o činnosti motora slúži ako spätná väzba pri kontrole správnosti priebehu meracieho procesu. Podľa zobrazenej informácie dokáže poučený operátor vhodne reagovať na možnú problémovú situáciu (potvrdiť chybu a pod.) Rovnaký textový zobrazovací prvok je použitý aj vo výsledkovej (3D Result) časti grafického prostredia.

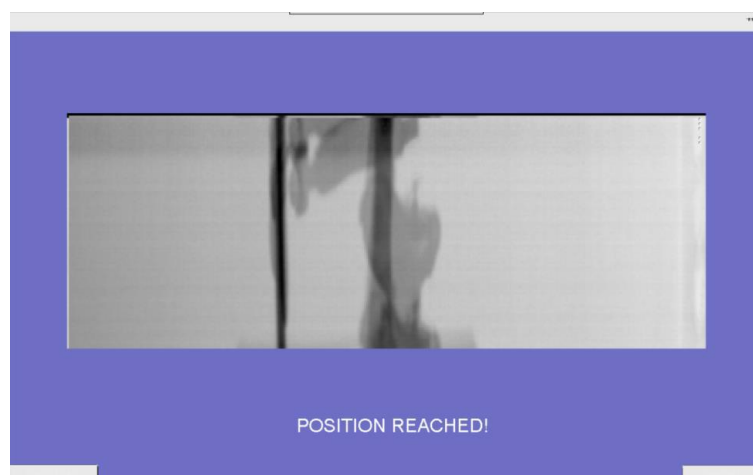
V časti MEASUREMENT užívateľ spúšťa automatizovaný proces merania – sekvenciu. V tejto časti je aj nastavenie referencie polohy, potvrdenie chyby a jednorazový posun motora pomocou príkazu MOVE POS a zadaním žiadanej polohy.



Obr. 3 Vizualizácia – časť ovládanie

Výsledková časť HMI zariadenia obsahuje Drawbox, v ktorom sú graficky zobrazené namerané údaje vo forme 3D obrázka. Tretím rozmerom obrázka je farebné rozlíšenie na škále od bielej po čiernu cez odtiene sivej. Aplikácia obsahuje 238 odtieňov, pomocou ktorých

zobrazuje konkrétnu vzdialenosť daného bodu od scanneru. Čím je objekt bližšie k scanneru, tým tmavšiu farbu mu algoritmus priradzuje.



Obr. 4 Vizualizácia – zobrazovacia časť 3D skenera

V. ZÁVER

V tejto práci bol predstavený návrh hardvérovej a softvérovej časti riešenia ovládania ochranného krytu 3D skenovacieho zariadenia. V rámci realizácie práce bola časť práce venovaná návrhu riadenia servopohonu pre polohovanie 3D skenera pomocou PLC a samotnej praktickej realizácie. V rámci budúcich prác na zariadení sa budú realizovať ďalšie doplňujúce funkcie pre prácu v sťažných podmienkach, ako je manažment ovládania ochranného priezoru a teplotný manažment.

POĎAKOVANIE

Táto práca bola podporovaná Agentúrou na podporu výskumu a vývoja na základe Zmluvy č. APVV-19-0210.

Táto práca vznikla s podporou grantu VEGA 1/0493/19.

LITERATÚRA

- [1] BUDIMIR, Miles: How to Select a PLC for Machine Control: An Engineer's Guide [online]. 2012. 1 s. [cit 2019-08-15] Dostupné na internete: <<https://www.motioncontroltips.com/choosing-a-plc-for-machine-control/>>
- [2] Assembly and Operating Manual [online].2018. 52 s. [cit 2019-08-15] Dostupné na internete: <<https://SCHUNK.com/fileadmin/pim/docs/IM0007121.PDF>>
- [3] Serial Communication [online]. 1 s.[cit 2020-02-25] Dostupné na internete: <<https://learn.sparkfun.com/tutorials/serial-communication/all#res>>.
- [4] Sectional diagram 0307300 PR 070-161-B [online].2019. 2 s. [cit 2020-03-03] Dostupné na internete: <https://SCHUNK.com/de_en/gripping-systems/series/pr/>
- [5] Motion Control SCHUNK V 1.60 [online].2015. 200 s. [cit 2020-03-03] Dostupné na internete: <<https://SCHUNK.com/fileadmin/pim/docs/IM0010976.PDF>>

Použitie počítačového videnia na ovládanie IoT zariadení pre hendikepovaných ľudí

¹Kristián MIČKO, ²Peter PAPCUN

¹Katedra Kybernetiky a Umelej Inteligencie, Fakulta Elektrotechniky a Informatiky
Technickej Univerzity v Košiciach, Slovenská Republika

²Katedra Kybernetiky a Umelej Inteligencie, Fakulta Elektrotechniky a Informatiky
Technickej Univerzity v Košiciach, Slovenská Republika

¹kristian.micko@tuke.sk, ²peter.papcun@tuke.sk

Abstrakt – Rozvoj viacerých disciplín informatiky umožnilo vytvoriť vhodné podmienky na tvorbu riešení problémov, s ktorými sa pravidelne stretávame na dennej báze. Vďaka pokroku v oblasti počítačového videnia, neurónových sietí, návrhu a výroby jednodoskových počítačov máme možnosť implementovať viaceré pomôcky v domácnosti, ktoré môžu pomôcť v jednotlivých jednoduchých úlohách hendikepovaným ľuďom v každodennom živote. Táto práca slúži ako ukážkou a demonštráciou toho, že kompozíciou viacerých predpripravených voľne dostupných knižníc a základných metód jednotlivých disciplín informatiky možno vyskladať fungujúci komplexný výstup na pomoc pri rôznych každodenných problémoch ľudí s rôznymi zdravotnými ťažkosťami. V súčasnej dobe sa vďaka spolupráci viacerých výskumných tímov a firiem podarilo vytvoriť toľko predpripravených konceptov, frameworkov a predtrénovaných modelov strojového učenia, že máme pred sebou výzvu aj hľadať uplatnenia týchto výstupov, aby sme z toho mohli vyťažiť maximum pre osoh rozvoja každodenného komfortu.

Kľúčové slová – Počítačové videnie, IoT, Tkinter, Python, OpenCV, Mediapipe, Raspberry Pi

I. ÚVOD

Ovládanie počítača za pomoci myši a klávesnice sa stalo nevyhnutnosťou každého osobného počítača. Na notebookoch sa namiesto myši používa touchpad. V súčasnej dobe sa používajú aj dotykové obrazovky na ovládanie jednotlivých zariadení ako napríklad tablet, smartphone či vybrané modely notebookov. Používanie dotykových displejov je na jednej strane pohodlné, ale na druhej strane má svoje nevýhody. Jedným z nevýhod je to, že zariadenie je ťažko ovládateľné alebo vôbec neovládateľné v prípade poškodenia displeja. Naopak zariadenia, ktoré používajú tlačidlá alebo klávesnicu a myš, vieme ovládať aj v prípade čiastočného zničenia obrazovky. V živote nastávajú aj také situácie, kedy potrebujeme z nejakého dôvodu ovládať výpočtové zariadenie mimo klasických prístupov vstupno-výstupných zariadení. Na tento štýl komunikácie môžeme použiť zvukové alebo vizuálne vnemy. Aplikácie vytvorené na ovládanie pomocou netradičného prístupu vstupno-výstupných zariadení vedia byť užitočné na kontrolu hendikepovaným ľuďom. Internet vecí je toho príkladom, kde jednotlivé zariadenia pripojené k elektrickej sieti dokážeme ovládať pomocou mikrokontrolerov/jednodoskových počítačov (SBC z ang. singleboard computer) pripojených na internet. SBC-ka väčšinou hostujú nejakú webovú, mobilnú alebo desktopovú aplikáciu na prevádzku služby diaľkového ovládania spotrebičov na internet. Tie služby si vyžadujú vo väčšine prípadov ovládanie cez tradičné zariadenia ako sú myši, klávesnice, touchpady alebo diaľkové ovládanie. Mnoho užitočných aplikácií možno implementovať za pomoci konceptu internetu vecí, ktoré môžu pomôcť hendikepovaným ľuďom. V obchode sú voľne dostupné senzory rôzneho druhu, ktoré môžu sa používať ako netradičný ovládací prvok počítača. Tento príspevok má za účel predviesť jednu z architektúr netradičného ovládania počítača. Využitie v praxi by mohlo priniesť to, aby napríklad ľudia s problémami so zrakom vedeli zapnúť nejaké spotrebiče v domácnosti bez rizika nejakého poškodenia vzhľadom na ich hendikép. Konkrétne, ak zasníma kamera slabozrakého prechádzať sa v nejakej miestnosti, aplikácia implementovaná za pomoci počítačového videnia by vedela usmerniť v miestnosti, kam sa má pohybovať, aby sa

predišlo zraneniu. Takisto aplikácia za pomoci hlasového asistenta môže pomôcť identifikovať daný spotrebič a ponúknuť mu možnosť jeho zapnutia alebo vypnutia s informáciou spotrebiča o jeho aktuálnom stave.

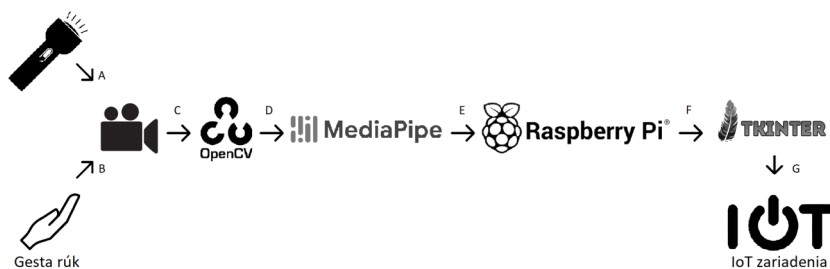
II. ANALÝZA SÚČASNÉHO STAVU

Technológie využívajúce elektroniku sa stali neoddeliteľnou súčasťou nášho života a pomáhajú nám riešiť každodenné situácie. Viaceré elektronické zariadenia boli v minulosti ovládané komplikovaným spôsobom a ich obsluha si vyžadovala prečítanie návodu na používanie. Postupom času sa začalo navrhovať ovládanie zariadení tak, aby to bolo, čo najviac intuitívne bežným používateľom, takzvané kognitívne riešenia. Začiatkom 70-tých rokov, kedy boli vyrobené prvé osobné počítače, sa začali skúmať spôsoby používateľský príjemného ovládania počítača [1]. Nárast domácnosti používajúce osobné počítače a inú elektroniku spôsobil postupnú automatizáciu aj v domácnostiach. Prvé smart domácnosti sa začali navrhovať z viacerých dôvodov ako napríklad pomoc pri opatere hendikepovaných ľudí začiatkom 70-tých rokov [2]. Rozvojom výpočtovej sily počítačov a objavením umelých neurónových sietí sa začali rozvíjať nové možnosti výskumu, vývoja a riešenie problémov, ktoré predtým bolo problém vyriešiť. Neurónové siete začali riešiť viaceré komplexné úlohy ako napríklad rozoznávanie tvári, gest rúk a podobne [3]. Predtrenované modely neurónových sietí môžeme použiť ako vstup do iných aplikácií. Dlhodobu sa navrhujú systémy na interakciu človeka s počítačom na základe gesta rúk, hlasu, emócií a iných verbálnych alebo neverbálnych spôsobov [4]. Takisto je možné v súčasnej dobe hrať videohry za pomoci virtuálnej reality (VR). VR a simulácia rôznych prostredí ako napríklad pilotovanie lietadla vedú aspoň čiastočne nahradiť fyzický hendikepovaným zážitky, ktoré by za normálnych okolností nemohli vyskúšať [5]. V prípade ľudí, čo majú problém s pamäťou, koncentráciou alebo s náhlými zmenami nálady vie byť nápocnou pomôckou aj inteligentné zrkadlo, ktoré môže ovládať viacero zariadení v domácnosti s pomocou hlasového asistenta Alexa od spoločnosti Amazon [6].

III. ARCHITEKTÚRA A IMPLEMENTÁCIA PROTOTYPU

A. Návrh architektúry

V tejto práci sme sa rozhodli použiť model smart osvetlenia a napájania ako ukážku implementovaného riešenia. Na Obr. 1 je znázornená schéma využitia technológií, ktoré sme sa rozhodli použiť na zhotovenie nášho prototypu. Náš prototyp sníma pozíciu rúk voči kamere a na základe predspracovania dát z kamery za pomoci knižnice *OpenCV* sa dáta pošlú do predtrenovaného modelu neurónovej siete na rozpoznávanie gest rúk. Táto neurónová sieť je prevádzkovaná knižnicou *mediapipe* a kam ruka ukáže na kamere, kde má ísť kurzor myši, tam sa myš ukáže na obrazovke. Ak blikne svetlo cez stred šošovky kamery riešenie bude to považovať ako kliknutie myši.



Obr. 1 Návrh architektúry implementácie riešenia z hľadiska softvérovej implementácie.

Rozhodnutie použiť bliknutie svetla cez stred šošovky bolo spôsobené tým, že samotná zmena gesta ruky zapríčinila neželaný pohyb kurzora myši na obrazovke ak chcel používateľ kliknúť. Na samotnú voľbu výberu zopínania/vypínania zásuvky sme používali *Tkinter* framework, ktorý ponúka ľahký spôsob návrhu a programovania GUI natívnymi aplikácií pod python a jednoduché prepojenie GPIO pinov pod Raspberry Pi 3b+ platformu.

B. Metodológia použitia počítačového videnia na pozadí implementácie riešenia

Každá digitálna kamera je postavená na princípe systému senzorov citlivých na určitú intenzitu a frekvenciu elektromagnetických vln, ktoré potom sú na výstupe obrazovky zobrazené v spektre viditeľného svetla. Počet fotocitlivých senzorov v matici a systém usporiadania optických šošoviek nám určuje maximálne rozlíšenie vytvorenej fotografie, ktoré môžeme vytvoriť pomocou danej

kamery [7]. Na rekonštrukciu farieb objektov sa najčastejšie používa model odtieňov červenej, modrej a zelenej známej pod skratkou RGB (z ang. red, green, blue), kde stupeň odtieňa sa určuje rozsahom hodnôt od 0 po 255. Táto vlastnosť reprezentácie obrázka nám v digitálnej podobe vytvára maticu v troch dimenziách, kde prvá dimenzia vraví o šírke, ďalšia o výške a posledná o kanáli odtieňoch z vybraných farieb RGB modelu [8].

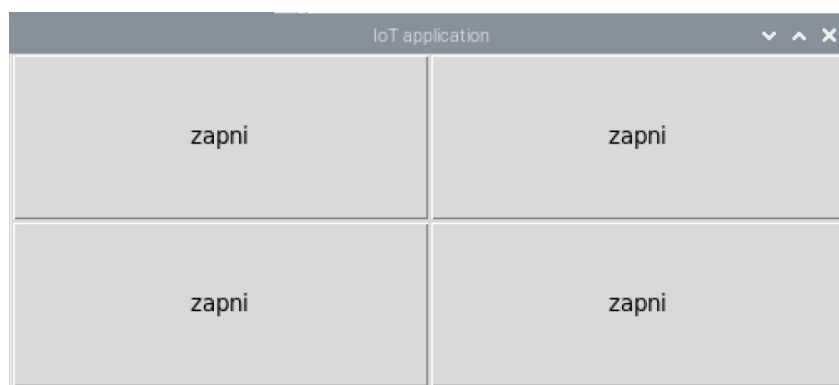
Vytvorenie maticovej reprezentácie obrazu opísanej spôsobom vyššie sa môže využiť rôznorodo na to, aby sme nad danou maticou robili rôzne matematické operácie, kde pomocou nich môžeme automaticky extrahovať informácie z obrázka a tak určovali ďalšie reakcie nejakého sofistikovanejšieho systému. Napríklad dáta reprezentované maticami sú ideálne ako vstup do výpočtu neurónovej siete, kde výstup neurónovej siete nám klasifikuje a lokalizuje nejaký cieľový objekt na danom obrázku [9].

V našom prípade sme v tejto práci použili túto znalosť na identifikáciu rúk snímaných kamerou. Neurónová sieť na pozadí nám identifikuje pozíciu ruky na živom vysielaní webkamery a podľa toho sa prepočíta relatívna pozícia kurzora myši na obrazovke.

Ako bolo spomenuté pri návrhu architektúry, kliknutie myši chceme simulovať bliknutím svetla cez stred šošovky kamery. Na tento účel si musíme uvedomiť, že RGB model aj keď je rozšírený, nie je jediným a na tento účel sme previedli obraz do modelu HSB (z ang. hue, saturation, brightness), čo predstavuje odtieň, sýtosť a jas v skratke. Tento model vie v porovnaní s RGB modelom lepšie rozpoznávať svetlo baterky od ostatných svetlých objektov na obrázku pri farebnej segmentácii.

C. Implementácia a ukážka prototypu

Na Obr. 2 možno vidieť grafické používateľské rozhranie implementované pre ovládanie modelu IoT zariadení použité v tejto práci.



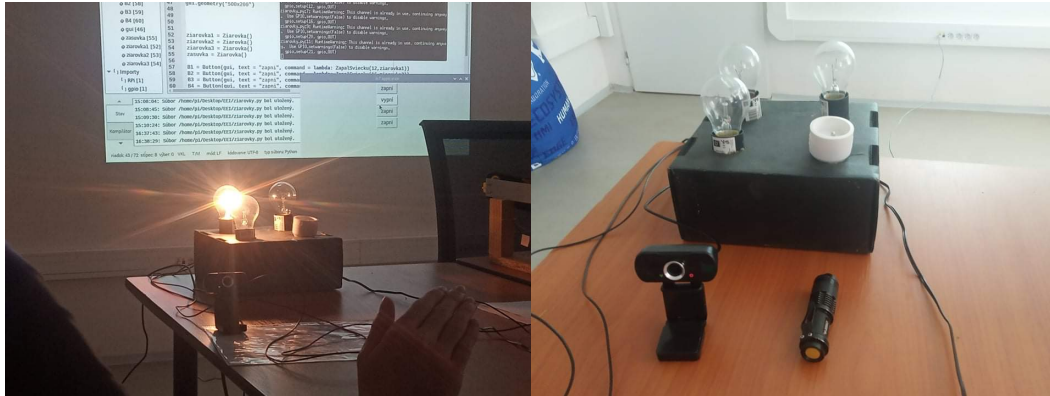
Obr. 2 GUI ovládania modelu na IoT veci vo framework-u tkinter.

Nápis zapni na tlačidlách sa automaticky zmení na vypni akonáhle zmeníme stav zapnutej resp. vypnutej zásuvky alebo žiarovky na použitom modeli. Tento minimalistický štýl grafického rozhrania nám zabezpečuje maximálnu intuitívnosť používania aplikácie.

Na Obr. 3 boli zobrazené ukážky fungovania prototypu na modeli IoT škatuľa, ktorá má pripojené 3 žiarovky a 1 zásuvku ovládanú za pomoci GPIO pinov napojených na Raspberry Pi 3B+. Ďalej je zobrazená webkamera s baterkou, ktorá na základe pohybov ruky a bliknutím cez stred šošovky kamery imitovalo používanie myši.

IV. ZÁVER A BUDÚCI VÝSKUM

Táto práca mala za účel poukázať na možnosti komunikácie medzi človekom a počítačom spôsobom viacej prirodzeným človeku za pomoci gest rúk. Na to nám poslúžili postupy počítačového videnia a predtrénovaného modelu rozpoznávania jednotlivých častí rúk. Aj napriek tomu, že ovládanie bolo prakticky bezkontaktné, čo umožňuje hendikepovaným ľuďom obsluhovať počítač bez toho, aby museli byť blízko počítača s myšou a klávesnicou, stále viacej dôležitou súčasťou napodobnenia komunikácie človeka s počítačom v štýle človek-človek je potrebné zahrnúť aj hlasovú komunikáciu. V súčasnej dobe sa pracuje aj na rozpoznávaní hovoreného slova a pochopenie významu textu za pomoci aplikácie neurónových sietí. Problémom u týchto systémoch je to, že každý človek na svete má svoj špecifický hlas, ktorý môže byť v niektorých prípadoch ťažko rozpoznateľný pre daný systém. Taktiež táto práca poukázala na to, že implementovaná komunikácia za pomoci gest rúk a počítačového videnia je použiteľná na jednodoskové počítače, čo umožňuje s ľahkosťou ovládať IoT veci v domácnosti ak človek z nejakých príčin sa nemôže



Obr. 3 Ukážka implementácie.

dostať k vypínačom dôležitých spotrebičov. Túto možnosť môžu veľmi oceniť hendikepovaní ľudia v každodennom živote.

V budúcom výskume sa chceme pozrieť na možnosť prepojenia prototypu na ovládanie za pomoci nielen počítačového videnia, ale aj rozpoznávania hovorených viet. Porovnať ich a vyhodnotiť použiteľnosť v praxi.

POĎAKOVANIE

Táto publikácia vznikla za podpory grantu robotické videnie v inteligentnom priestore (FEI-2022-88).

LITERATÚRA

- [1] F. E. Ritter, G. D. Baxter, and E. F. Churchill, "User-centered systems design: a brief history," in *Foundations for designing user-centered systems*. Springer, 2014, pp. 33–54.
- [2] J. Bennett, O. Rokas, and L. Chen, "Healthcare in the smart home: A study of past, present and future," *Sustainability*, vol. 9, no. 5, p. 840, 2017.
- [3] Y.-c. Wu and J.-w. Feng, "Development and application of artificial neural network," *Wireless Personal Communications*, vol. 102, no. 2, pp. 1645–1656, 2018.
- [4] R. P. Sharma and G. K. Verma, "Human computer interaction using hand gesture," in *ELEVENTH INTERNATIONAL CONFERENCE ON COMMUNICATION NETWORKS, ICCN 2015/INDIA ELEVENTH INTERNATIONAL CONFERENCE ON DATA MINING AND WAREHOUSING, ICDMW 2015/INDIA ELEVENTH INTERNATIONAL CONFERENCE ON IMAGE AND SIGNAL PROCESSING, ICISP 2015*, ser. Procedia Computer Science, P. Shenoy, S. Iyengar, K. Raja, K. Venugopal, R. Buyya, and L. Patnaik, Eds., vol. 54. Bangalore Univ, Univ Visvesvaraya Coll Engn; Govt India, Tech Educ Qual Improvement Programme II; Alpha Coll Engn; Vijaya Vittala Inst Technol; Soc Informat Proc, 2015, pp. 721–727, 11th International Conference on Image and Signal Processing (ICISP), Bangalore, INDIA, AUG 21-23, 2015.
- [5] D. Lee, K. Baek, J. Lee, and H. Lim, "A development of virtual reality game utilizing kinect, oculus rift and smartphone," *International Journal of Applied Engineering Research*, vol. 11, no. 2, pp. 829–833, 2016.
- [6] P. S. K. Swamy, B. JSSATEB, S. India Aditya, S. A. Ashish, G. Hemanth Reddy, and S. A. Jeevan, "Smart mirror with alexa using raspberry pi."
- [7] K. R. Spring, "Scientific imaging with digital cameras," *Biotechniques*, vol. 29, no. 1, pp. 70–76, 2000.
- [8] T. Porter and T. Duff, "Compositing digital images," in *Proceedings of the 11th annual conference on Computer graphics and interactive techniques*, 1984, pp. 253–259.
- [9] J. Du, "Understanding of object detection based on cnn family and yolo," in *Journal of Physics: Conference Series*, vol. 1004, no. 1. IOP Publishing, 2018, p. 012029.

Bezsnímačové riadenie synchrónneho motora s permanentnými magnetmi v oblasti nízkych rýchlostí

Viktor Petro

Katedra elektrotechniky a mechatroniky, Fakulta elektrotechniky a informatiky,
Technická univerzita v Košiciach, Slovenská republika.

viktor.petro@tuke.sk

Abstrakt – Bezsnímačové riadenie synchrónneho motora s permanentnými magnetmi (SMPM) je atraktívna oblasť výskumu z dôvodu možnosti úspory na cene, ako aj zvýšenia bezpečnosti pohonu. Robustný a precízny pozorovateľ polohy rotora je nevyhnutný pre implementáciu bezsnímačového riadenia. V oblasti nízkych rýchlostí sa najčastejšie využívajú metódy založené na injektáži vysokofrekvenčného (VF) signálu do statorového vinutia. Hlavná nevýhoda bežne používaných metód je použitie digitálnych filtrov v prúdovej slučke, resp. v procese estimácie polohy a uhlovej rýchlosti. Tieto digitálne filtre nepriaznivo pôsobia na pásmo priepustnosti regulátorov. V tomto článku bude uvedená jednoduchá, ale efektívna metóda bezsnímačového riadenia, v ktorej sú digitálne filtre eliminované. Simulačné výsledky potvrdzujú vysokú dynamiku bezsnímačového riadenia.

Keywords – Synchrónny motor s permanentnými magnetmi, bezsnímačové riadenie, injektáž vysokofrekvenčného signálu, pozorovateľ

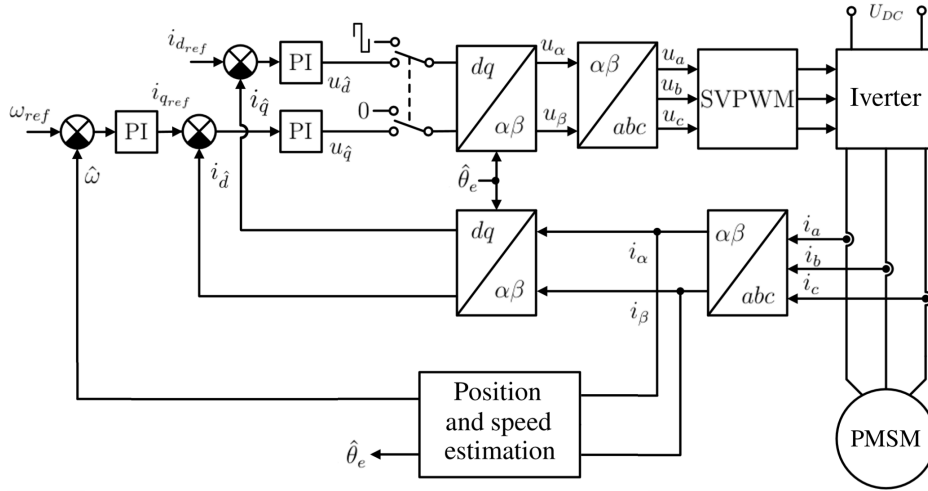
I. ÚVOD

Synchrónny motor s permanentnými magnetmi je často používaným motorom v priemysle kvôli vysokej účinnosti, dynamickým charakteristikám a bezúdržbosti. Pre riadenie motora je obvykle adoptované vektorové riadenie, ktoré poskytuje kvalitnú reguláciu v celej pracovnej oblasti stroja. Avšak pre tento typ riadenia je nevyhnutná presná informácia o polohe rotora voči statorovému vinutiu. Pre tento účel sa používajú mechanické snímače polohy, ako inkrementálny snímač, resolver a pod. Nevýhodou je, že tieto snímače negatívne vplyvajú na celkovú robustnosť stroja, nakoľko sú citlivé na elektromagnetické rušenie, okrem toho navyšujú celkové náklady na pohon, vyžadujú montážny priestor atď. Výhody eliminácie snímača polohy sú známe a preto je táto oblasť predmetom výskumu v posledných desaťročiach. Tieto metódy sú spoločne označované ako bezsnímačové riadenie [1].

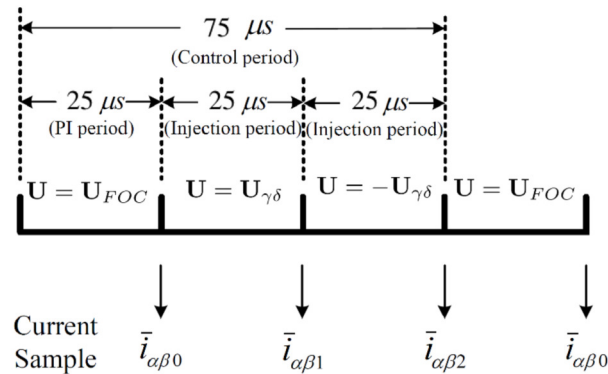
Vo všeobecnosti platí, že existujúce metódy bezsnímačového riadenia je možné aplikovať pre oblasť vysokých, alebo pre oblasť nízkych rýchlostí. Metódy pre vysoké rýchlosti sú väčšinou založené na pozorovaní indukovaného napätia. Keďže amplitúda indukovaného napätia je úmerná uhlovej rýchlosti rotora, tieto metódy zlyhávajú v oblasti nízkych až nulových rýchlostí [2]. Pre oblasť nízkych rýchlostí boli vyvinuté metódy založené na VF injektáži signálov do statorového vinutia stroja. Pri týchto metódach sa využíva magnetická nesymetria stroja, spôsobená geometrickými vlastnosťami motora, resp. presýtením magnetického obvodu. Injektovaný signál môže mať rôznu frekvenciu a tvar. Najviac používané typy injektovaných signálov sú:

- injektáž VF signálu so sínusovým priebehom,
 - injektáž v statorovom súradnicovom systéme [3],
 - injektáž v rotorovom súradnicovom systéme [4],
- injektáž VF signálu so obdĺžnikovým priebehom [5],
- injektáž pulzujúceho VF signálu [6].

Historicky prvým prístupom boli metódy založené na injektáži VF signálu so sínusovým priebehom. Hlavná nevýhoda spočíva v obmedzení frekvencie injektovaného signálu, pre zachovanie sínusového tvaru. Vylepšením je injektáž VF signálu s obdĺžnikovým priebehom. V oboch prípadoch sa injektáž vykonáva súčasne s vektorovým riadením a VF signál je superponovaný na základnú frekvenciu riadiaceho napätia. Následne prúdy tiež obsahujú okrem základnej



Obr. 1 Bloková schéma bezsnímačového riadenia použitím injektáže pulzujúceho VF signálu.



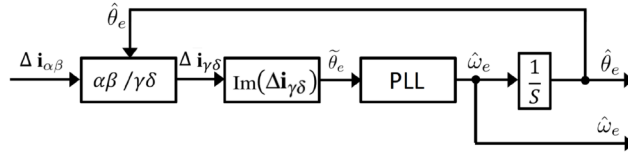
Obr. 2 Riadiaca perióda bezsnímačového riadenia.

frekvencie aj VF zložky, ktoré je potrebné odseparovať. Docieli sa to použitím digitálnych filtrov, pričom zložka zo základnou frekvenciou sa použije ako spätná väzba pre prúdové regulátory a VF zložka sa použije pri estimácii rýchlosti a polohy rotora. Digitálne filtre nepriaznivo pôsobia na pásmo priepustnosti regulátor, následkom čoho je zhoršená dynamika pozorovania ako aj celkového bezsnímačového riadenia motora. V súčasnosti sa výskum zaoberá metódami, pri ktorých je injektovaný signál oddelený od riadenia motora. Následne nie sú potrebné digitálne filtre v procese estimácie ani v spätných väzbách regulátorov a tieto metódy vykazujú vyššiu dynamiku a lepšie vlastnosti bezsnímačového riadenia motora. Do tejto skupiny patrí napr. injektáž pulzujúceho VF signálu, ktorý bude priblížený v nasledujúcej kapitole.

II. INJEKTÁŽ PULZUJÚCEHO VF SIGNÁLU

Bloková schéma bezsnímačového riadenia pomocou injektáže pulzujúceho VF signálu je zobrazená na Obr. 1. Prepínač v blokovej schéme prezentuje fakt, že buď sa vykonáva vektorové riadenie, alebo sa vykonáva injektáž. Celková riadiaca perióda je tvorená tromi spínacími periódami, ako je to zobrazené na Obr. 2. Spínacia frekvencia je zvolená ako 40 kHz (25 μ s) a celková riadiaca perióda bola vykonávaná frekvenciou 13.33 kHz (75 μ s). V prvej spínacej perióde je vykonávané vektorové riadenie pričom PI regulátory prúdov vytvoria napäťový vektor \mathbf{U} . V nasledujúcich dvoch spínacích periódách je vykonaná injektáž kladnej a zápornej polvlny injektovaného signálu. V rámci týchto dvoch injektážnych periódách nie je vektorové riadenie vykonávané (resp. je prerušené). Injektáž je vykonávaná v pozorovanej dq rotorovej súradnicovej sústave, označovaná aj ako $\gamma\delta$ referenčná sústava. Prúd je meraný na začiatku každej spínacej periódy, ako to je zobrazené na Obr. 2. Pre prúdové zmeny medzi dvomi spínacími periódami v $\alpha\beta$ referenčnej súradnicovej sústave platí:

$$\Delta \mathbf{i}_{\alpha\beta} = (c_1 + c_2 e^{j2(\theta_e - \theta_u)}) \mathbf{u}_{\alpha\beta} \Delta T, \quad (1)$$



Obr. 3 Výpočet pozorovanej uhlovej rýchlosti a elektrickej polohy rotora z nameraných prúdových zmien.

kde $c_1 = \frac{\Sigma L}{\Sigma L^2 + \Delta L^2}$, $c_2 = \frac{-\Delta L}{\Sigma L^2 + \Delta L^2}$, kde $\Sigma L = \frac{L_d + L_q}{2}$, $\Delta L = \frac{L_d - L_q}{2}$ a L_d, L_q sú indukčnosti v d a q osi, θ_e je skutočná poloha rotora, θ_u je uhol natočenia napäťového vektora v $\alpha\beta$ súradnicovej sústave a $\mathbf{u}_{\alpha\beta} = U_m e^{j\theta_u}$, kde U_m je amplitúda napäťového vektora $\mathbf{u}_{\alpha\beta}$.

Pretože injektáž je vykonávaná v pozorovanej $\hat{d}\hat{q}$ ($\gamma\delta$) sústave, je výhodné transformovať (1) do pozorovanej rotorovej súradnicovej sústavy. Pre tento účel sa zavedie symbol $\hat{\theta}_e$ označujúci pozorovanú polohu rotora [6]:

$$\Delta \mathbf{i}_{\gamma\delta} = \Delta \mathbf{i}_{\alpha\beta} e^{-j\hat{\theta}_e} = (c_1 + c_2 e^{j2(\theta_e - \hat{\theta}_e - \hat{\theta}_u)}) \mathbf{u}_{\gamma\delta} \Delta T, \quad (2)$$

kde $\mathbf{u}_{\gamma\delta}$ je napäťový vektor v pozorovanej rotorovej súradnicovej sústave $\hat{d}\hat{q}$ ($\gamma\delta$), $\hat{\theta}_u = \theta_u - \theta_e$ je uhol natočenia napäťového vektora v pozorovanej súradnicovej sústave a ΔT spínacia perióda. Zložka prúdu v d osi nevytvára moment a preto je injektáž vykonávaná v pozorovanej \hat{d} (γ) osi. Pre injektované kladné a záporné napäťové pulzy platí:

$$\begin{aligned} \mathbf{u}_{\gamma\delta 1} &= U_m e^{j\hat{\theta}_{u1}} \approx U_m e^{j0} - \Delta \mathbf{u}, \\ \mathbf{u}_{\gamma\delta 2} &= U_m e^{j\hat{\theta}_{u2}} \approx U_m e^{j\pi} - \Delta \mathbf{u}, \end{aligned} \quad (3)$$

kde $\mathbf{u}_{\gamma\delta 1}$, $\mathbf{u}_{\gamma\delta 2}$ predstavuje kladnú a zápornú injektážnu periódu, $\hat{\theta}_{u1}, \hat{\theta}_{u2}$ sú uhly injektovaných napäťových vektorov v pozorovanej $\hat{d}\hat{q}$ ($\gamma\delta$) sústave a $\Delta \mathbf{u}$ predstavuje napäťové úbytky spôsobené nelinearitou striedača. Dosadením (3) do (2) dostaneme:

$$\begin{aligned} \Delta \mathbf{i}_{\gamma\delta 1} &= \Delta T c_1 (U_m e^{j0} - \Delta \mathbf{u}) + \Delta T c_2 e^{j2\hat{\theta}_e} e^{-j\hat{\theta}_{u1}} U_m, \\ \Delta \mathbf{i}_{\gamma\delta 2} &= \Delta T c_1 (U_m e^{j\pi} - \Delta \mathbf{u}) + \Delta T c_2 e^{j2\hat{\theta}_e} e^{-j\hat{\theta}_{u2}} U_m, \end{aligned} \quad (4)$$

kde $\tilde{\theta}_e = \theta_e - \hat{\theta}_e$ predstavuje chybu v pozorovanej polohe rotora. Pretože injektáž je vykonávaná v reálnej osi, chyba v pozorovanej polohe je úmerná prúdovým zmenám v imaginárnej osi nasledovne:

$$\text{Im}(\Delta \mathbf{i}_{\gamma\delta 1} - \Delta \mathbf{i}_{\gamma\delta 2}) = 2k \sin(2\tilde{\theta}_e) \approx 4k\tilde{\theta}_e, \quad (5)$$

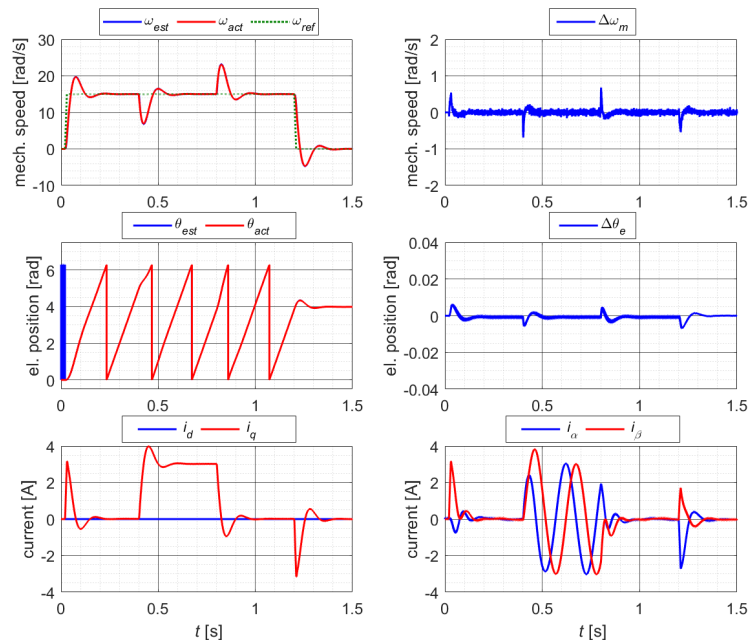
kde $k = \Delta T c_2 U_m$ a pre jednotlivé prúdové zmeny počas injektážnych periód $\Delta \mathbf{i}_{\gamma\delta 1}$, $\Delta \mathbf{i}_{\gamma\delta 2}$ platí:

$$\begin{aligned} \text{Im}(\Delta \mathbf{i}_{\gamma\delta 1}) &= -(i_{\alpha 1} - i_{\alpha 0}) \sin \hat{\theta}_e + (i_{\beta 1} - i_{\beta 0}) \cos \hat{\theta}_e, \\ \text{Im}(\Delta \mathbf{i}_{\gamma\delta 2}) &= -(i_{\alpha 2} - i_{\alpha 1}) \sin \hat{\theta}_e + (i_{\beta 2} - i_{\beta 1}) \cos \hat{\theta}_e. \end{aligned} \quad (6)$$

Pozorovanú uhlovú rýchlosť $\hat{\omega}_e$ a polohu rotora $\hat{\theta}_e$ je možné získať fázovým závesom (phase-lock loop - PLL) podľa Obr. 3. Najprv sa vypočítajú prúdové zmeny v $\alpha\beta$ súradnicovej sústave počas dvoch injektážnych periód. Následne je použitá Parková transformácia pre transformáciu prúdových zmien do pozorovanej rotorovej súradnicovej sústavy $\hat{d}\hat{q}$ ($\gamma\delta$). Počas transformácie je použitá pozorovaná poloha $\hat{\theta}_e$. Následne sa použije len imaginárna zložka prúdových zmien, keďže tá nesie informáciu o polohe rotora. Všetky tieto tri kroky sú zahrnuté v (6). Ak chyba v pozorovanej polohe $\hat{\theta}_e$ je dotlačená do nuly, pozorované veličiny konvergujú ku skutočným hodnotám a pozorovaná rotorová súradnicová sústava $\hat{d}\hat{q}$ ($\gamma\delta$) sa zrovná so skutočnou dq sústavou.

III. SIMULAČNÉ VÝSLEDKY

Simulácia bezsnímačového riadenia bola vykonaná pomocou softvéru MATLAB/Simulink. Nominálne napätie medziobvodu pre použitý motor je $U_{DC} = 230$ V, indukčnosti $L_d = 0.012$ mH, $L_q = 0.034$ mH a odpor vinutia $R_{ab} = 6.98$ Ω . Spínacia frekvencia bola nastavená na 40 kHz. Amplitúda injektovaného napätia bola zvolená ako $U_m = 40$ V. Simulačné výsledky sú zobrazené na Obr. 4. Referenčné napätie bolo nastavené na 15 rad/s (143 rpm). Počas prechodových dejov bola maximálna chyba v pozorovanej polohe $\Delta\theta_e$ okolo 0.006 rad (0.34 elektrických stupňov) a maximálna chyba v pozorovanej uhlovej rýchlosti bola okolo 0.5 rad/s (5 rpm). V čase $t = 0.4$ s bol motor zaťažovaný nominálnym momentom.



Obr. 4 Simulačné výsledky bezsnímačového riadenia.

IV. ZÁVER

V tomto článku bola uvedená metóda bezsnímačového riadenia SMPM založená na injektážii VF pulzujúceho napätia. Oddelením injektáže a vektorového riadenia motora sa docieli eliminácia digitálnych filtrov, čo má za následok celkové zlepšenie dynamiky a vlastnosti bezsnímačového riadenia. Simulačné výsledky potvrdzujú presnosť, vysokú dynamiku a vhodnosť použitia tejto metódy pre bezsnímačové riadenie.

POĎAKOVANIE

Táto práca bola podporená Fakultou elektrotechniky a informatiky, Technickou univerzitou v Košiciach, pod grantom: Grant FEI-2022-86 Experimentálne overenie bezsnímačového riadenia SMPM s integrovaným meničom.

REFERENCES

- [1] Y. Hosogaya and H. Kubota, "Position estimating method of ipmsm at low speed region using dq-axis current derivative without high frequency component," in *2013 IEEE 10th International Conference on Power Electronics and Drive Systems (PEDS)*, 2013, pp. 1306–1311.
- [2] V. Petro and K. Kyslan, "Design and simulation of direct and indirect back emf sliding mode observer for sensorless control of pmsm," *Power Electronics and Drives*, vol. 5 (40), pp. 215–228, 2020. [Online]. Available: <https://doi.org/10.2478/pead-2020-0016>
- [3] P. Jansen and R. Lorenz, "Transducerless position and velocity estimation in induction and salient ac machines," *IEEE Transactions on Industry Applications*, vol. 31, no. 2, pp. 240–247, 1995.
- [4] M. Corley and R. Lorenz, "Rotor position and velocity estimation for a permanent magnet synchronous machine at standstill and high speeds," in *IAS '96. Conference Record of the 1996 IEEE Industry Applications Conference Thirty-First IAS Annual Meeting*, vol. 1, 1996, pp. 36–41 vol.1.
- [5] R. Ni, K. Lu, F. Blaabjerg, and D. Xu, "A comparative study on pulse sinusoidal high frequency voltage injection and inform methods for pmsm position sensorless control," in *IECON 2016 - 42nd Annual Conference of the IEEE Industrial Electronics Society*, 2016, pp. 2600–2605.
- [6] G. Xie, K. Lu, S. K. Dwivedi, J. R. Rosholm, and F. Blaabjerg, "Minimum-voltage vector injection method for sensorless control of pmsm for low-speed operations," *IEEE Transactions on Power Electronics*, vol. 31, no. 2, pp. 1785–1794, 2016.
- [7] D. Calvetti and L. Reichel, "On the evaluation of polynomial coefficients," *Numerical Algorithms*, vol. 33, pp. 153–161, Aug. 2003.
- [8] V. Y. Pan, "Solving a polynomial equation: some history and recent progress," *SIAM Review*, vol. 39, pp. 187–220, June 1997.
- [9] E. M. Johansson and J. E. Mast, "Three-dimensional ground-penetrating radar imaging using synthetic aperture time-domain focusing," *Advanced Microwave and Millimeter Wave Detectors, Proceedings of SPIE*, vol. 2275, pp. 205–214, Sep. 1994.

Sledovanie Parametrov Fotovoltaických Článkov

¹Marek PAVLÍK, ²Dávid DZÚR

^{1,2} Department of Electric Power Engineering, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Slovak Republic

¹marek.pavlik@tuke.sk, ²david.dzur@student.tuke.sk

Abstract — Tento príspevok pojednáva o zmene parametra sériového rezistora. V článku je popísaný náhradný model fotovoltaického článku, ktorého súčasťou je aj sériový rezistor. Jeho zmena ovplyvňuje celkovú výrobu z fotovoltaických článkov a následne panelov. V príspevku sa popisuje aj vplyv sériového odporu na V-A charakteristiku fotovoltaických článkov zapojených v sérii. Z výsledkov vyplýva, že zvýšenie hodnoty sériového odporu ovplyvňuje tvar V-A charakteristiky a výkonovej charakteristiky.

Keywords — fotovoltaický článok, MPP bod, napätie naprázdno, prúd nakrátko, string

I. ÚVOD

Slnko je zdrojom obnoviteľnej energie, ktoré je zložené prevažne z dvoch prvkov: z vodíka 73,46 % a hélia 24,85 %. Vyžarovanie energie v podobe elektromagnetického žiarenia prebieha v obrovských množstvách 24 hodín denne 365 dní v roku. Energia slnka je s ľudstvom od vzniku samotného života na planéte Zem. Ľudia sa časom naučili slnečnú energiu využívať vo svoj prospech od ohrievania vody, sušenia oblečenia a prípravy jedla až po zmenu tejto energie na elektrinu.

Slnko vyžaruje každú sekundu energiu $3,8 \cdot 10^{23}$ kW. Z toho množstva žiarenia dopadne na planétu Zem, ktorá je vzdialená od Slnka 150 miliónov km, len $1,8 \cdot 10^{14}$ kW. 60 % žiarenia dopadá na zemský povrch a zvyšných 40% je odrazených naspäť do kozmu a pohltých v atmosfére.

Poznáme tri typy slnečného žiarenia – Obr.1: [1] [2]

- Priame žiarenie: žiarenie Slnka, ktoré priamo dopadá na povrch bez rozptylu
- Difúzne žiarenie: žiarenie rozptýlené prechodom cez mraky a nečistoty v atmosfére
- Odrazené žiarenie: odrazené od objektov a povrchu zeme



Obr.. 1 Typy slnečného žiarenia – 1. priame, 2. difúzne, 3. odrazené [1]

Celkové žiarenie je tvorené súčtom priameho, difúzneho a odrazeného žiarenia, dopadajúceho na horizontálny povrch.

$$I_C = I_P + I_D + I_O \quad (1)$$

Silné stránky fotovoltaickej produkcie elektrickej energie sú najmä významný technický potenciál, decentralizovaná výroba v mieste spotreby napr. fotovoltaika na streche domu, zníženie objemu emisií skleníkových plynov a minimálne prevádzkové náklady. K nevýhodám môžeme zaradiť nízku elektrickú účinnosť, výroba elektrickej energie je premenlivá a závisí od ročného obdobia. Pri fotovoltaických paneloch je to dokonca denná variabilita závislá od počasia a oblačnosti v oblasti inštalácie. Pre bežnú populáciu je azda najväčšou prekážkou relatívne vysoká náročnosť prvotnej finančnej investície a dlhšie obdobie návratnosti investície. [2]

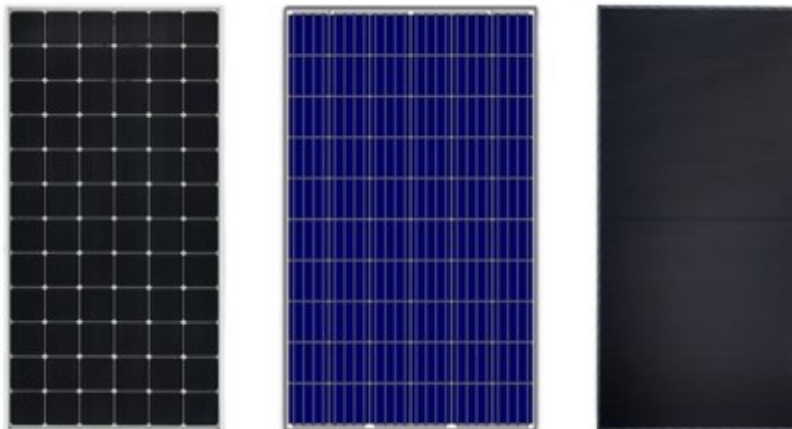
II. DRUHY FOTOVOLTAICKÝCH ČLÁNKOV

Fotovoltaické články sú vyrábané zo svetlo citlivého polovodiča, ktorý vplyvom fotónov uvoľňuje elektróny potrebné pre vznik jednosmerného elektrického prúdu. Typy fotovoltaických článkov rozdeľujeme do dvoch hlavných celkov: fotovoltaické články z kryštalického kremíka a tenko vrstvomé fotovoltaické články. [2]

Monokryštalické články (Obr.2) dosahujú účinnosť premeny svetelnej energie na elektrickú energiu 13 % až 17 %, sú široko rozšírené a pri vhodných svetelných to sú najefektívnejšie solárne články. Tento typ článku dokáže premeniť žiarenie 1000 W/m² na 140 W elektrickej energie pri ploche článku 1 m². Produkcia monokryštálov vyžaduje absolútne čistý polovodičový materiál. Očakávaná životnosť je 25-30 rokov, časom však rovnako ako všetky fotovoltaické články degradujú. Panely tvorené monokryštálom sú rozpoznateľné podľa článkov tvaru štvorcov so skosenými rohmi, skosenie je zapríčinené optimalizáciou tvaru pri rezaní z kruhového prierezu čím vzniká vyššie množstvo nevyužitého kremíka. Farba článkov je čierna. V roku 2020 predstavovali monokryštalické články 80 % celkovej produkcie fotovoltaických článkov. [2]

Polykryštalické články (Obr.2) sú ekonomickejšie a menej náročné na výrobu, tekutý kremík sa vyleje do foriem tvaru kvádrov, pri procese tuhnutia sa vytvárajú kryštalové štruktúry rôznych veľkostí, na hraniciach týchto štruktúr však môžu vznikať nežiadúce defekty, ktoré znižujú celkovú účinnosť daného článku. Účinnosť sa pohybuje v intervale od 10 % do 14 % so životnosťou medzi 20 až 25 rokov. Výhodou polykryštalických článkov je menší pokles výkonu pri difúznom žiarení a preto sú vhodnejšie pre inštaláciu do oblastí s častým výskytom oblačnosti. [2]

Tenkovrstvomé články (Obr.2) sú vytvárané vrstvením polovodičového materiálu hrúbky 0,3 až 2 mikrometre na sklenený alebo plastový substrát prípadne substrát z nerezovej oceli. Náklady sú pre minimálnu hrúbku materiálov veľmi nízke. Nevyužívajú sa na inštaláciu v domácnostiach pre ich nízku účinnosť v porovnaní s kryštalickými článkami. Najznámejšie typy tenkovrstvomých článkov sú: CdTe (kadmium, telúr), CIGS (meď, indium, gálium, selén), GaAs (gálium, arzén) a tenkovrstvomé kremíkové panely, ktoré sa následne delia na: amorfné panely (a-Si), amorfné tandemové články (mikromorf) a tenkovrstvomý polykryštalický kremík v skle. [2] [3]

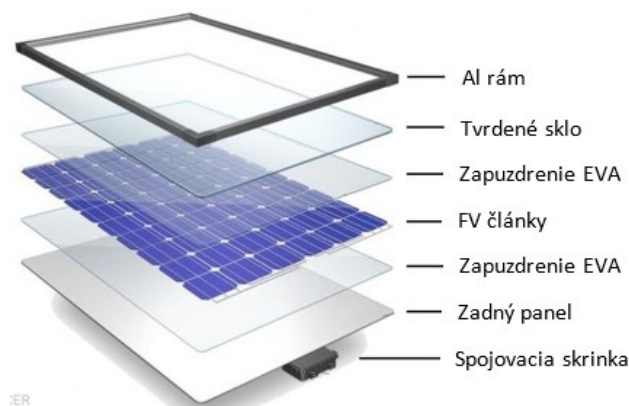


Obr. 2 Fotovoltaické panely – monokryštalický, polykryštalický a tenkovrstvomý fotovoltaický panel [2]

Samotný fotovoltaický článok generuje len málo elektrickej energie. Na zvýšenie výkonu sa články spájajú do skupín nazývaných solárne moduly resp. panely. Solárny panel je tvorený hliníkovým rámom, v ktorom je osadené tvrdené sklo hrúbky 2-4 mm slúžiace na ochranu fotovoltaických článkov pred mechanickými účinkami prostredia napr. krupobitie. Pod tvrdeným sklom sa nachádzajú samotné články zapuzdrené v EVA (etylenvinylaretát) fólii, ktorá chráni najmä pred vlhkosťou a vonkajšími vplyvmi prostredia. Panel je zo zadnej strany uzavretý krytom

najčastejšie z PVF, PET alebo PP, na ktorý sa pripevňuje spojovacia skrinka využívaná na elektrické prepojenie panelov, skrinka je vybavená by-pass diódami. Pri zatičení alebo znečistení panela musíme ochrániť panel pred spätným prúdom, usmernenie toku prúdu len jedným smerom zabezpečuje by-pass dióda. [2]

Vplyvom zatičenia fotovoltaického článku dochádza k zníženiu výkonu, pri fotovoltaických článkoch zapojených do série je výsledný prúd stringu rovnaký ako prúd generovaný zatičeným fotovoltaickým článkom, v prípade celkového zatičenia článku je výsledný prúd stringu rovný 0 A. Zatičený článok sa správa ako záťaž, čím sa mení smer prúdu a tieneny článok sa zahrieva – vznikajú teplé miesta tkz. hotspot. Fotovoltaické články sú v solárnom paneli zapojené v sérii, časť celkového množstva fotovoltaických článkov zapojených v sérii je substring. Ku každému substringu je paralelne zapojená bypass dióda. Zatičenie je najčastejšie spôsobované iným objektom, lístím, prachom alebo snehom. Bypass diódou pri bežnej prevádzke netečie žiadny prúd, v prípade zatičenia sa dióda otvára a elektrický prúd obteká postihnutú časť stringu. [2]



Obr. 3 Stavba fotovoltaického panela [2]

III. MOŽNOSTI ZAPOJENIA FOTOVOLTAICKÉHO SYSTÉMU

Fotovoltaické systémy rozdeľujeme podľa zapojenia na 3 skupiny a to [2][4][5]:

1. Fotovoltaický systém pripojený k distribučnej sieti (on-grid)
2. Fotovoltaický systém nepripojený k distribučnej sieti (off-grid)
3. Hybridný systém

Systém on-grid predstavuje prepojenie fotovoltaického systému s distribučnou elektrickou sieťou. Vyrobena elektrická energia sa dodáva do distribučnej siete. V tomto prípade zapojenia na sieť nie sú potrebné akumulátory, čím sa znižujú náklady na výstavbu systému. Hlavnými časťami zapojenia sú solárne panely, striedač a elektromer a distribučná sieť.

Ostrovná prevádzka (off-grid) sa najčastejšie využíva v oblastiach bez možnosti pripojenia k elektrickej sieti, napr. odľahlá chata v lese alebo v horách. Pri takejto prevádzke sme priamo odkázaní len na elektrickú energiu zo solárnych panelov a z vopred nabitých akumulátorov. Ostrovný systém je nákladnejší pre potrebu akumulátorov, ktoré majú nižšiu životnosť ako samotné solárne panely. Takéto zapojenie si vyžaduje regulátor nabíjania a podľa potreby aj napäťový striedač. Problémové ročné obdobie je zima, kedy slnečný svit trvá najkratšie, ďalší kritický problém je sneženie, ktoré pokrýva a tieni panely. Ostrovné prevádzky majú predpoklad byť hlavným pilierom, ktorý pomôže elektrifikovať vidiecke oblasti v krajinách tretieho sveta.

Hybridný systém spája výhody systému pripojeného k sieti a ostrovného systému. Takýto systém môže byť vybavený akumulátorom a zároveň je pripojený k distribučnej sieti. Akumulátory poskytujú záložný zdroj elektrickej energie a pripojenie k sieti umožňuje dodávanie energie do siete v čase prebytku. Hlavnou výhodou je možnosť napájania objektu aj pri výpadku elektrickej siete, zapojenie dokáže napájať aj spotrebiče s väčším odberom. Cena je vyššia ako on-grid systém a odhadovaná životnosť akumulátora sa pohybuje od 7-15 rokov. [2] [6]

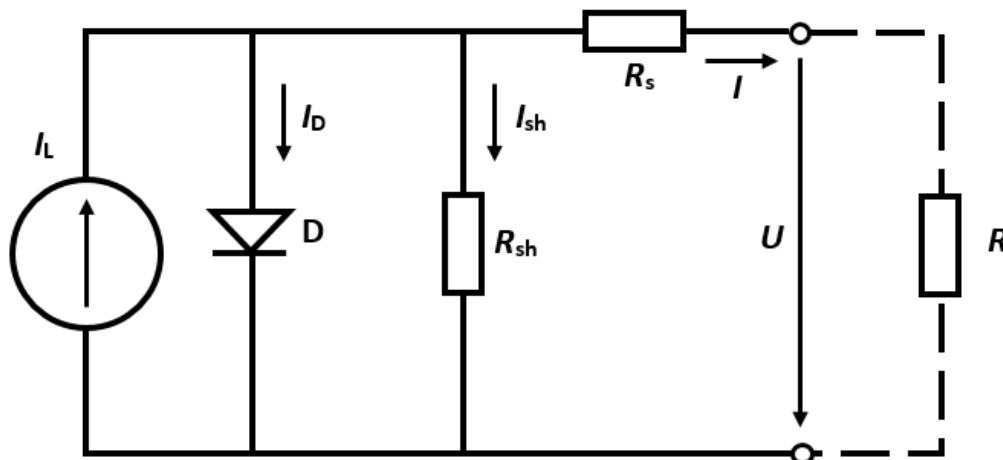
Pre správne pochopenie fungovania fotovoltaického článku je vhodné vytvoriť ekvivalentný model, ktorý bude zložený z prvkov, ktorých správanie je už známe. Náhradná schéma fotovoltaického článku pozostáva z prúdového zdroja s paralelne pripojenou diódou a rezistorom, na modelovanie prechodového odporu sa využíva odpor zapojený do série.

Prúd dodávaný fotovoltaickým článkom do záťaže je sformulovaný podľa Kirchoffovho zákona o prúdoch: [2]

$$I = I_L - I_D - I_{sh} \text{ [A]} \quad (1)$$

Úplná formulácia prúdu fotovoltaického článku je reprezentovaná rovnicou: [2]

$$I = I_L - I_S \cdot \left[\exp\left(\frac{U+IR_S}{nU_T}\right) - 1 \right] - \frac{U+IR_S}{R_{sh}} \quad [A] \quad (2)$$



Obr. 4 Náhradná schéma fotovoltaického článku

V rovnici prúd I_L reprezentuje prúd generovaný žiarením vo fotovoltaickom článku, I_D je prúd tečúci diódou a prúd I_{sh} je stratový prúdom prechádzajúcim paralelným odporom R_{sh} . Paralelný odpor je spôsobený miestnymi defektmi PN priečodu na okrajoch fotovoltaických článkov, ideálna hodnota tohto odporu je nekonečne veľká. Vo vysoko kvalitných článkoch je $R_{sh} = 300 - 400 \text{ k}\Omega$. [2]

V prípade náhradnej schémy s jednou diódou je prúd I_D vyjadrený pomocou Shockleyho rovnice pre ideálnu diódu: [2]

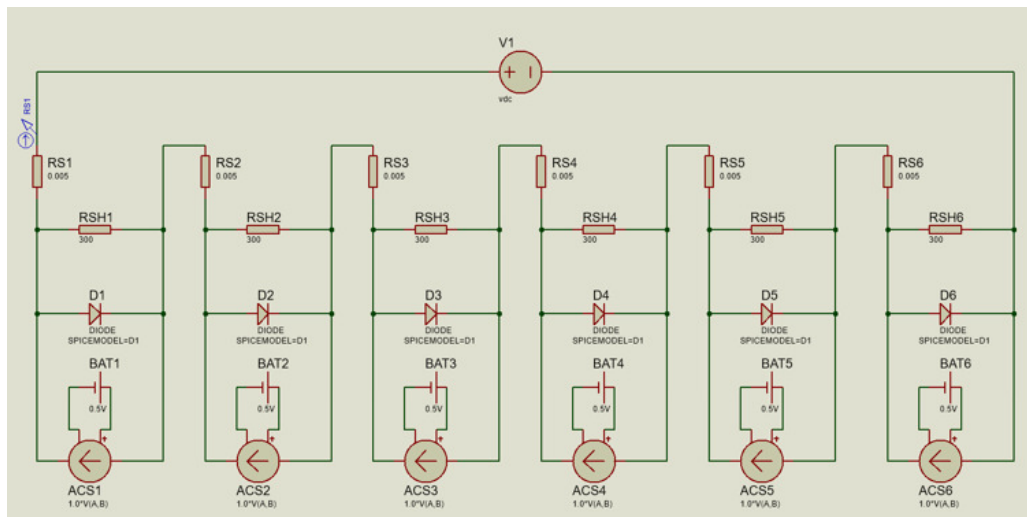
$$I_D = I_S \cdot \left[\exp\left(\frac{U+IR_S}{nU_T}\right) - 1 \right] \quad [A] \quad (3)$$

kde I_D je prúd na dióde, I_S je saturačný prúd, U_D je napätie na dióde, q ($= 1.6 \times 10^{-19} \text{ C}$) je elektrónový náboj, k ($= 1.38 \times 10^{-23} \text{ J/K}$) je Boltzmannova konštanta a T je termodynamická teplota pri ktorej daná dióda pracuje.

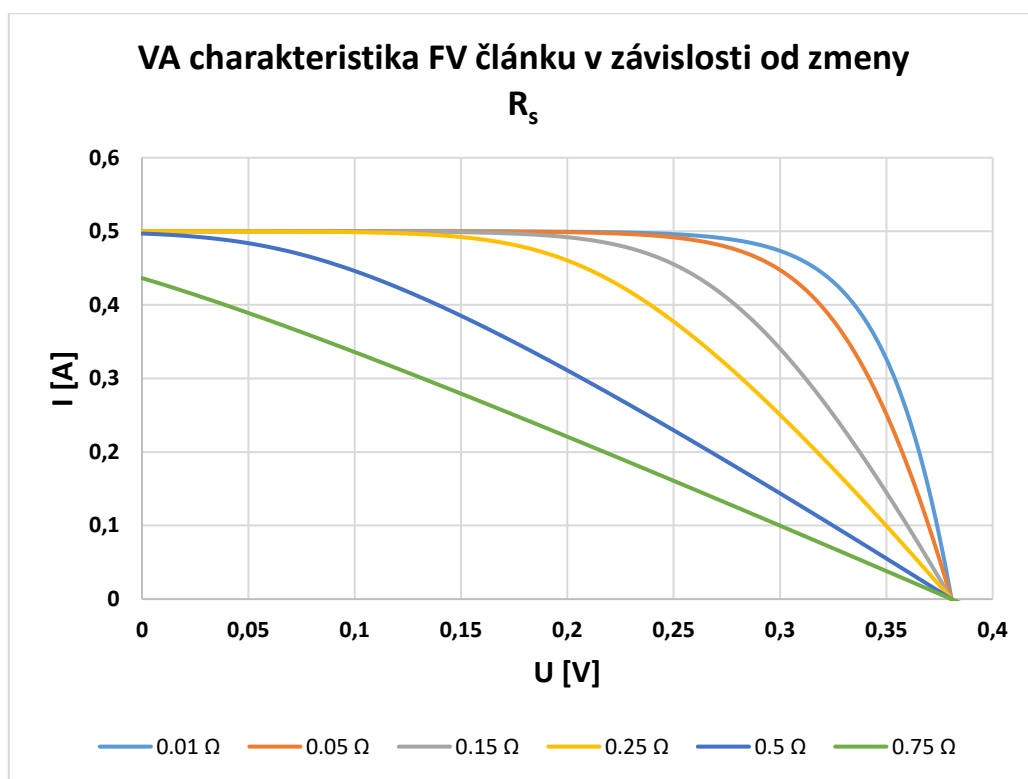
Odpor R_S je sériový odpor, hodnota tohto odporu by mala byť čo najnižšia, tento odpor je zapríčinený nedokonalosťami spojov medzi jednotlivými článkami. U kvalitných článkov je veľkosť odporu R_S do $0,10 \Omega$. Veľičina n udáva faktor ideálnosti diódy, hodnota sa pohybuje v intervale $1 - 2$.

IV. PREŠETRENIE ZMENY PARAMETROV FOTOVOLTAICKÉHO ČLÁNKU – ZMENA SÉRIOVÉHO ODPORU R_S

Simulácia parametrov FV článkov bolo namodelované v programe protheus. Schéma zapojenia je zobrazená na Obr. 5. Po zapojení prvkov a nastavení parametrov ekvivalentnej schémy fotovoltaického článku v programe Proteus, prebehla zmena parametra sériového odporu R_S , ktorý vo fotovoltaických článkoch predstavuje odpor vodičov a prechodový odpor. Zvýšenie tohto odporu spôsobuje pokles výkonu fotovoltaického článku. Zo simulácie vyplýva, že so zvyšovaním sériového odporu fotovoltaického článku dochádza k poklesu elektrického prúdu tečúceho záťažou. Dodávaný prúd do záťaže pri zvýšenom sériovom odpore fotovoltaického článku klesá. Bod maximálneho výkonu MPP, ktorý sa nachádza v kolene charakteristiky sa so zvyšujúcim sériovým odporom fotovoltaického článku začína vyskytovať už pri nižšej hodnote napätia. Pri vyšších hodnotách odporu R_S sa oblasť kolena úplne stráca a voltampérová charakteristika (Obr. 6) nadobúda tvar klesajúcej priamky. Po zvýšení sériového odporu simulovaného článku na hodnotu $R_S = 0,75 \Omega$ došlo už aj k výraznému poklesu prúdu nakrátko. Klesá faktor plnenia aj výkon z FV článku. Intenzita slnečného žiarenia bola počas každej simulácie konštantná $E = 1000 \text{ W/m}^2$, paralelný odpor článku bol taktiež konštantný $R_{sh} = 30 \text{ k}\Omega$.

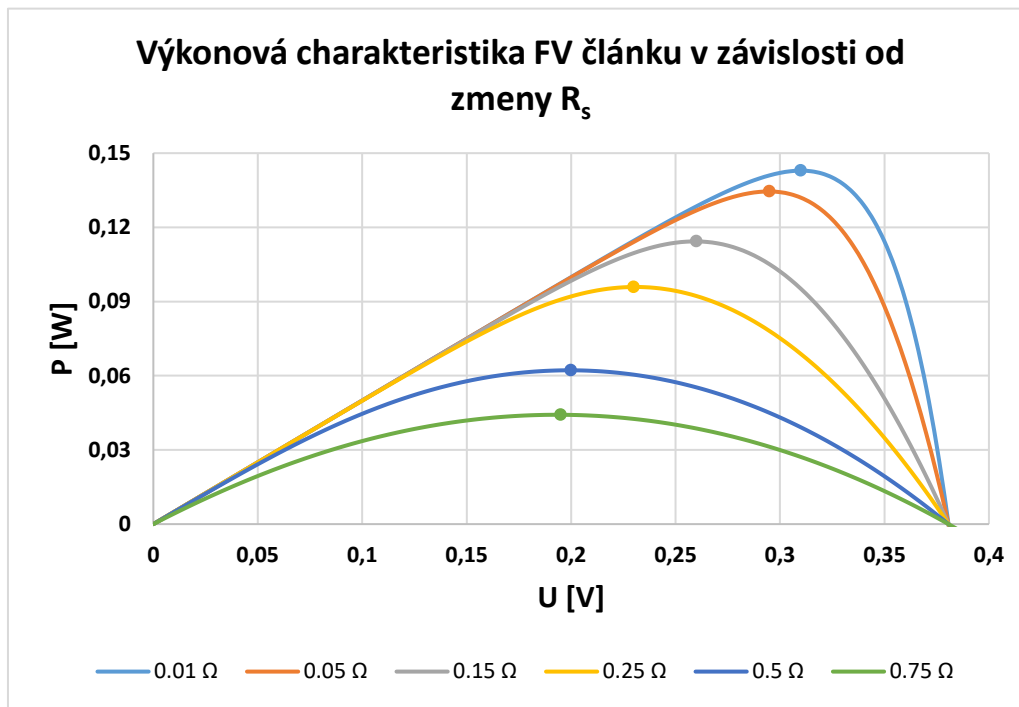


Obr. 5 String tvorený 6ks FV článkov

Obr. 6 V-A charakteristika FV článkov pri zmene R_s

Vo výkonovej charakteristike FV článku (Obr.7) je vidieť negatívny vplyv zmeny sériového odporu na výkon článku. Výkon článku pri zvýšenom sériovom odpore fotovoltaického článku výrazne klesá a súčasne dochádza k rozširovaniu oblasti, v ktorej by sa mohol bod MPP nachádzať. S nárastom R_s dochádza vo výkonovej charakteristike ku skracovaniu lineárnej časti, pri vysokých hodnotách sa lineárna časť stráca a výkonová charakteristika nadobúda tvar oblúka.

Skratový prúd simulovaného obvodu sa pri sériovom odpore do hodnoty 0,25 Ω nemenil a mal konštantnú hodnotu 0,50 A.. Napätie naprázdno bolo v každej simulácii konštantné, $U_{OC} = 0,38$ V. Maximálny výkon článku mal hodnotu $P_{MAX} = 0,143$ W, pri prúde $I_{MAX} = 0,461$ A a napätí $U_{MAX} = 0,310$ V. Zvýšenie hodnoty sériového odporu znižuje faktor plnenia, ktorý bol maximálny pri najmenej simulovanej hodnote odporu $R_s = 0,01$ Ω , faktor plnenia mal v tomto prípade hodnotu 0,752 resp. 75,2 %.



Obr. 7 Výkonová charakteristika FV článkov pri zmene R_s

V. ZÁVER

Tento článok pojednáva o vyšetrení sériového odporu R_s . Bol vytvorený model fotovoltaických článkov, ktorého súčasťou je sériový rezistor s hodnotou odporu. Tento rezistor predstavuje sériové straty dané nedokonalosťou spojov a taktiež prechodový odpor medzi fotovoltaickými článkami.

Z výsledkov simulácie vyplýva, že zmena hodnoty sériového rezistora ovplyvní parametre fotovoltaických článkov. Táto zmena sa prejaví aj v prípade fotovoltaických panelov a stringov. Zo simulácie vyplýva, že so zvyšovaním sériového odporu fotovoltaického článku dochádza k poklesu elektrického prúdu tečúceho záťažou. Dodávaný prúd do záťaže pri zvýšenom sériovom odpore fotovoltaického článku klesá. Bod maximálneho výkonu MPP, ktorý sa nachádza v kolene charakteristiky sa so zvyšujúcim sériovým odporom fotovoltaického článku začína vyskytovať už pri nižšej hodnote napätia. Pri vyšších hodnotách odporu R_s sa oblasť kolena úplne stráca a voltampérová charakteristika (Obr. 6) nadobúda tvar klesajúcej priamky.

POĎAKOVANIE

Táto práca bola podporená Agentúrou na podporu výskumu a vývoja na základe zmluvy č. APVV-19-0576, ako aj Vedeckou grantovou agentúrou Ministerstva školstva, vedy, výskumu a športu SR a SAV, podporená grantom VEGA 1/0757/21.

LITERATÚRA

- [1] D. Dzúr, „Sledovanie parametrov fotovoltaických článkov,“ Bakalárska práca, 2022.
- [2] M. Pavlík, „Obnoviteľné zdroje energie vo všeobecnosti,“ 1.vyd. Technická univerzita v Košiciach, 2019. 75 s. ISBN 978-80-553-3317-5.
- [3] J. Zhang, et. al., "Model Parameter Analysis of Cracked Photovoltaic Module under Outdoor Conditions," 2020 47th IEEE Photovoltaic Specialists Conference (PVSC), 2020, pp. 2509-2512, doi: 10.1109/PVSC45281.2020.9300720.
- [4] N. Das, et.al., "Modelling and analysis of multi-junction solar cells to improve the conversion efficiency of photovoltaic systems," 2014 Australasian Universities Power Engineering Conference (AUPEC), 2014, pp. 1-5, doi: 10.1109/AUPEC.2014.6966482.
- [5] T. Maniak, "Modelovanie parametrov fotovoltaických článkov," bakalárska práca, 2021, Technická univerzita v Košiciach, pp.77.
- [6] D. Gupta, et.al, "Photovoltaic Modeling using Single Diode Model in MATLAB," 2020 IEEE International Conference on Computing, Power and Communication Technologies (GUCON), 2020, pp. 734-739, doi: 10.1109/GUCON48875.2020.9231165.

Local currency creation using blockchain technology

¹Radoslav FIGURA, ²Branislav MADOSŤ, ³Anton BALÁŽ

^{1,2,3}Faculty of Electrical Engineering and Informatics, Department of computers and informatics, Technical University of Košice, Slovak Republic

¹radoslav.figura@student.tuke.sk, ²branislav.mados@tuke.sk, ³anton.balaz@tuke.sk

Abstract – This paper explains the concepts and implementation strategies for developing digital local cryptocurrency with a complete ecosystem. The proposed implementation of local currency utilizes Ethereum blockchain, a decentralized P2P platform technology. The implementation of local currency is implemented and deployed using smart contracts, which communicate with a web application to provide basic operations for the user.

Keywords – blockchain, cryptocurrencies, programming, local economics, local currency, complementary currency

I. INTRODUCTION

Currently we observe multiple states of financial systems development. Governments, banks and society have unified value to simple value bearing tokens such as bills, banknotes and coins in contrast to the past when more often some sort of bartering systems were predominant. The primary issue with bartering was the inability to devise some bearers of value (could not just cut cows in half). Modern sophisticated financial systems utilize some form of consensus to agree that a certain medium of exchange has a value. This allowed for the implementation of digital currency and digital accounts. According to figures from the European Central Bank, there are around 1.6 trillion euros in physical cash [1] compared to over 20 trillion euros in digital form [2]. More consumers are paying for products and services with digital currency and influencing which forms of payment retailers accept. Digitization provides several advantages in terms of security, transparency, options, and access, and society develops new forms of payment based on these advantages. Financial systems constantly develop, and so do concepts of complementary money. This paper is focused on providing conceptual ideas for digital complementary currencies that are distinct to any geographical region, as well as describing and proposing potential solutions for the deployment of local currencies utilizing smart contracts.

II. PARALLEL MONEY

This section introduces concepts of parallel money. Discussed concepts of parallel money are applied in III. section for conceptual model of local currency. This section provides a brief overview of digital cryptocurrencies as a platform for parallel money.

A. Complementary currencies

Complementary currencies describe two schemes for exchange of goods, services and knowledge between small groups of subjects using their own currency parallel to fiat currency. First scheme that is most often used describes coupons, shop credits and benefits. Another type of complementary currency mentioned in this paper is self-governed currency with structures, institutions and members that is very similar to fiat currency which is maintained by central banks and governing entities. Huge advantage of complementary currency is it being democratic, small, closed, and adaptive [3]. In small communities anyone can suggest what could be changed and if rules are not set in favor of all members no one will use that currency neither will anyone accept it. There are still many ongoing debates about usage of parallel currencies, but they are of no relevance globally [4]. Their primary objective is to temporarily boost sellers, production, region development and people [5].

B. Cryptocurrencies

Cryptocurrencies broadly speaking represent new forms of digital payment gateways and are a complex financial institution that provides a way to pay, maintain, communicate, provide and use digital tokens bearing value. There are numerous cryptocurrencies but the vast majority is built on blockchain technology. Blockchain technology uses P2P communication between all communicating nodes providing system resources and data. Data in blockchain are stored in blocks of data, written and read using transactions, and are assigned to accounts [6]. Multiple blockchains allow adjusting accounts in such ways that they could act as a program, these are called smart contracts [7]. Suggested solution of local currency uses smart contracts programs that communicate with other accounts on ethereum blockchain using Ether as cryptocurrency.

III. CONCEPTS OF LOCAL CURRENCIES

Based on the previous section, new local currency as a concept can be modeled. This section provides advanced sophisticated concepts of investing, bonuses, fees, governing bodies, members, and roles for the system to be functional and self-maintaining.

A. Local currency digitization

To implement currency there must be a way of exchanging that currency between parties and members that accept it. These exchange systems introduce payments and transfers.

There are a few mechanics of local currencies that must be met as well. Complementary currency must provide a means to obtain usable funds or other forms of tokens acceptable by a group of members. There are multiple types of those tokens like coupons, coins, papers, bills, time [8], vocal agreement, and digital data which is the preferred token in digitization. Acquiring usable digital data is possible by exchanging fiat currency. In addition to being able to exchange complementary tokens for fiat currency, it should be feasible to withdraw funds from parallel currency. Local currencies are also geographically restricted [3]. The suggested concept of checking the payer's current location whilst paying or transferring money to another account should ensure payments are made locally without the presence of a third party.

Local currency concepts can presuppose that there is a definite real entity (local government, institution, executive administration) responsible for the maintenance of local currency. If people believe, embrace, and accept local currency rules and concepts they can become members of the local currency.

B. Full ecosystem

For an ecosystem to function, all entities are accountable for maintaining local money. For people to join a local currency, there must be an incentive such as a bonus for a token deposit. These bonus tokens cannot be created arbitrarily and they must be represented by real value such as fiat money. Fiat money dedicated for bonuses are provided by investors, who invest with their whole portfolio. Anyone can become an investor by investing any amount of fiat money. Investors are compensated for their service when customers withdraw funds or transfer tokens.

For users to engage with this system, they must be able to spend their tokens. Utilizing services and purchasing things from local companies is the most evident way to spend tokens. The system provides a way for businesses to mark their location on a digital map for mainly marketing purposes and a better user experience.

To ensure money circulation and prevent stagnation of user accounts, a mechanism that punishes these activities as a concept was proposed. Users whose accounts remain inactive for more than five years are immediately terminated. Their accounts are considered defunct, and the owner taxes their tokens at a rate of 100 percent. On the other hand, when accounts are inactive for more than 3 years, the owner-defined stagnation tax is charged.

The owner is responsible for maintenance and local currency activity. Every owner should define a set of regulations that are beneficial and advantageous to all parties as a motivational factor. The owner cannot maintain members. The whole ecosystem is maintained automatically by the system outlined in the conceptual model.

IV. SOLUTION

In this section recommended concepts of the local currency are described in conjunction with a suggested solution. Each subsection delves into concepts and explains them in greater detail.

A. Implementation

Implementation of the local currency as a smart contract was achieved using the ethereum blockchain environment infrastructure. Program and instructions for data writing were achieved with a program written using a solidity programming language in Remix IDE. The user application is used to facilitate access to any ethereum account, display the account's balance, give a method for obtaining funds, communicate with any local currency, deploy new local currency, and maintain local currency. The application for communication with blockchain is entirely maintained by a system administrator, who is also responsible for the deployment of smart contracts, the availability of a private ethereum blockchain, and the enforcement of the local currency stagnation policy.

B. Maintenance, taxation and tax revenue

Many internal operations for local currency are initiated by the owner. The owner is responsible for establishing taxes on withdrawal of money, and token transfer and establishes taxes for inactive accounts stagnation. The owner determines how many bonus tokens are awarded when a token is deposited. The owner determines the distribution of revenue from transfer tax between investors, owner, and system administrator. The owner can alter a defined region by expanding or contracting it, and can also define a new owner.

C. Token transfer

Any entity can use token transfer to move funds from one account (payer) to another (beneficiary). The transfer is only possible if the paying account has sufficient tokens and is within the local currency's defined region. Transfer tax is applied if the conditions are met, and the final amount is credited to the beneficiary's account as shown on the fig. 1.

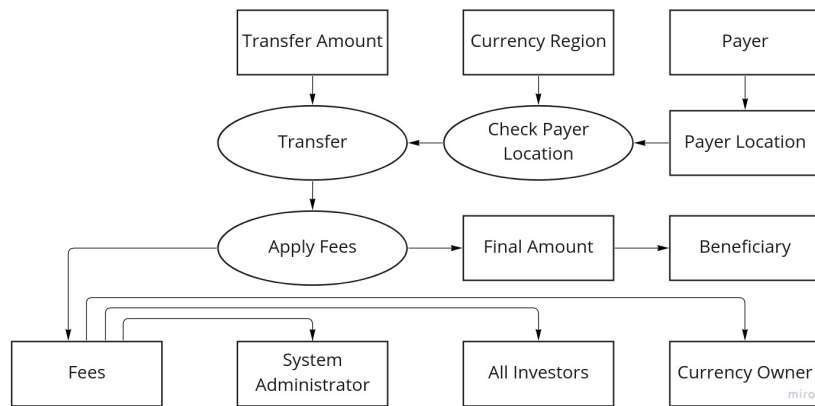


Fig. 1 Transfer token model and fees

D. Investing for bonuses

Investing guarantees that token depositors will receive bonus tokens. The allocation of bonus tokens is an active process that could be described as a competition between investors. The system attempts to select one investor for token allocation. The investor is selected based on his portfolio state, his current account balance, and his tenure as an investor. The oldest investors are given priority. To perform bonus allocation, only investors with sufficient funds and the highest ratio of initially invested fiat money to account balance are chosen. By allocating a greater number of bonuses investors have a higher chance of winning when users withdraw money from the local currency. When withdrawing funds, the user is required to pay withdrawal taxes, which are then allocated to the most recent investor who has the greatest number of successfully allocated bonuses.

E. Stagnation

Stagnation introduces anti-hoarding prevention. The system attempts to penalize users who do not actively manage their accounts by purchasing and transferring tokens. There is a Javascript application in the infrastructure that communicates directly with the ethereum blockchain that controls and stagnates all users with inactive accounts using CRON jobs every day at noon by communicating with all local currencies. When inactive user accounts are detected, the system

chooses between two tax types as shown on fig. 2. 100% dead tax is applied to inactive accounts that have been inactive for more than five years. Other tax is defined by the owner of local currency and is applied to user accounts that are not active for more than 3 years. The tax revenue is then added to the owner's portfolio. As it is the responsibility of the owner to maintain and provide resources for the local currency, the owner should use these tokens in favor of the defined geographical area.

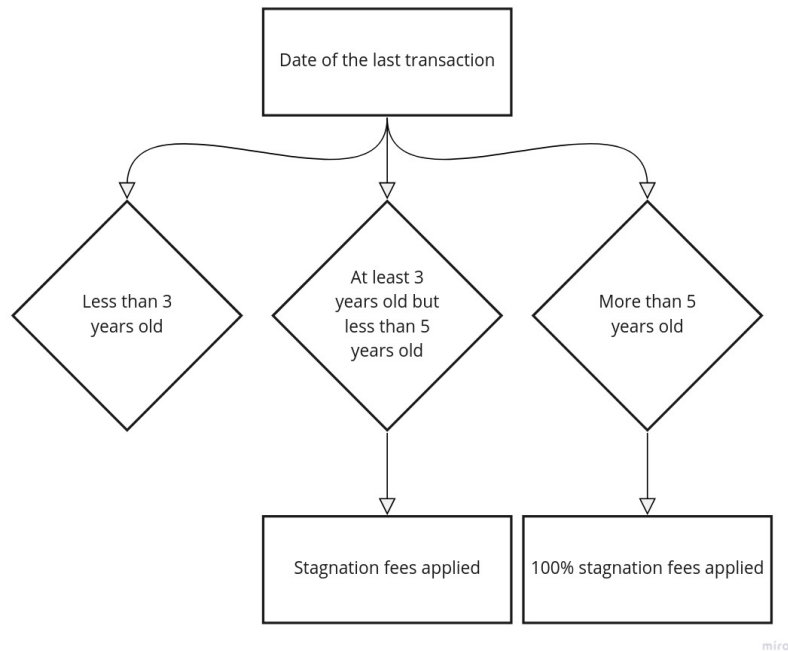


Fig. 2 Stagnation fees model

V. USER INTERFACE

The user interface allows users to communicate, use, maintain and interact with local currency. This section explains the local currency dashboard, which serves as a direct interface for communicating with individual local currencies, as well as the automated deployer, which is used to create new local currencies and deploy them as smart contracts on the ethereum private blockchain.

A. Local currency dashboard

User interface (fig. 3) is developed by meeting designed concepts. It is designed to be intuitive, convey sufficient information, and facilitate user interactions in as few steps as possible. The user can transfer tokens, deposit tokens, and withdraw money. These actions are fundamental actions for users to perform. An information about user account balance, and user's roles in local currency are displayed to user. All users have access to advanced interactions, such as creating businesses, maintaining businesses, and investing money. The owner has the ability to alter taxes, revenues, and the local currency owner. The owner can also recreate the region of influence. All events and issues are logged to the user. The user can also examine all available investors.

B. Automated deploy and system for local currency overview

The user interface provides a way for users to create new local currency free of charge (fig. 4) by defining area of influence, local currency name and local currency symbol. An application for communication with blockchain then takes provided information and populate provided information into template of smart contract, compiles smart contract and deploys smart contract to private ethereum blockchain with which the application communicates. When a smart contract is deployed all necessary information are stored in conventional MongoDB NoSQL database. All steps and logs are available to the new local currency owner, who created the local currency. The new local currency is available in list of all local currencies. If a user accepts the local currency's rules and configuration, they are free to join any formed local currency.

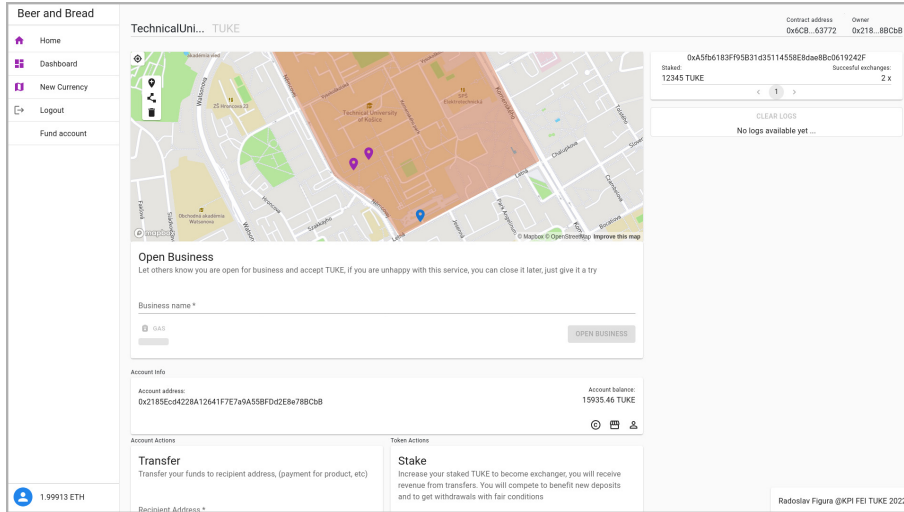


Fig. 3 User interface of web application

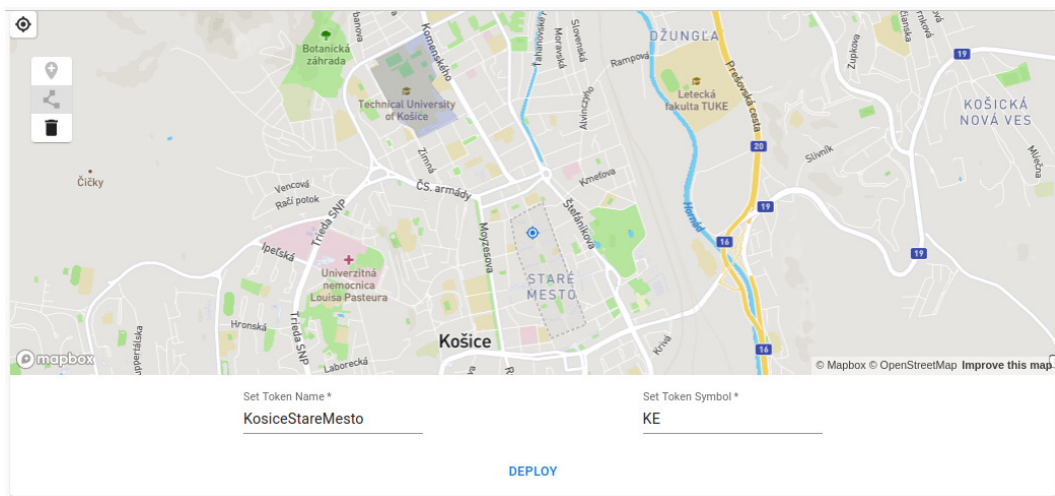


Fig. 4 Deployment of new local currency

VI. RESULTS

Testing a solution that has been implemented is a way to see whether system works and is both usable and useful. In this section comparison with public blockchain was performed. In order to ensure a good user experience, the transaction speed was also examined.

A. Testnet and private blockchain

In the proposed approach, a private ethereum blockchain was used. It has numerous advantages over the public blockchain in terms of proposed solutions and entity hierarchies. Private blockchains are superior in terms of speed prices for transactions which are significantly lower, but not in terms of decentralization. Ropsten testnet was used in comparison as public blockchain in order to compare solely the times of actions displayed in the table 1.

Action	Private (times in ms)	Public (times in ms)
Depositing funds	525	29209
Investing	697	63877
Tax Maintenance	285	62155
Transfer	528	27156

Table 1
 Comparison testnet and private blockchain

B. Concurrent transactions

To test multiple transactions and execution times for all transactions to be deemed successful, different transaction amount was initiated with each call, and the execution time required to complete them was recorded. The transaction procedure includes verifying the user’s location, redistributing revenue from transaction taxes, and updating the user’s most recent transaction information. All of these processes contributed to the final execution time of bulks of transactions. The table 2 displays information on transaction times.

Table 2
 Execution time needed for transaction completion

Amount of transactions	Time of execution (ms)
1	81
2	89
4	171
8	211
16	458
32	768
64	1778
128	3284
256	6926
512	13574
1024	30261
2048	65867
4096	167678
8192	520935
16384	990234

VII. CONCLUSION

In conclusion, the proposed approach meets the requirements to be called a local currency. Description of full ecosystem implies that there are methods of verifying user location, providing access to tokens bearing value, and withdrawing funds from closed economics. Advanced approaches proposed portrayed the system as an environment that can actively self-maintain with minimal to no external influence. The implemented user interface enables all required activities to be performed from a single, concise interface, hence enhancing the user experience.

ACKNOWLEDGMENTS

This publication has been published with the support of the Operational Program Integrated Infrastructure within project: Research in the SANET Network and Possibilities of Its Further Use and Development (ITMS code: 313011W988), co-financed by the ERDF. This research was supported by KEGA 002TUKE-4/2021 Implementation of Modern Methods and Education Forms in the Area of Cybersecurity Towards Requirements of Labour Market.

REFERENCES

- [1] E. C. Bank, “Banknotes and coins circulation,” Jan 2021. [Online]. Available: https://www.ecb.europa.eu/stats/policy_and_exchange_rates/banknotes+coins/circulation/html/index.en.html
- [2] —, “Monetary statistics,” April 2022. [Online]. Available: https://sdw.ecb.europa.eu/quickview.do?SERIES_KEY=117.BSLA.U2.N.L.LE0.A.1.Z5.0000.Z01.
- [3] M. S. Cato and N. Dodd, *Rebirth of an old technology*. New Economics Foundation, 2015, p. 31–41.
- [4] D. Liberto, “Community currency definition,” 9 2021. [Online]. Available: https://www.investopedia.com/terms/c/community_currencies.asp
- [5] M. Naqvi and J. Southgate, “Banknotes, local currencies and central bank objectives - the history of alternative currency schemes,” *Quarterly Bulletin*, p. 1–9, 2013.
- [6] S. Nakamoto, “Bitcoin: A peer-to-peer electronic cash system,” *Decentralized Business Review*, pp. 1–9, 2008.
- [7] R. Cordell, “Introduction to smart contracts,” 10 2020. [Online]. Available: <https://ethereum.org/en/developers/docs/smart-contracts/>
- [8] E. Cahn, “Time banking: An idea whose time has come?” *Yes! Magazine*, 11 2011. [Online]. Available: <https://www.yesmagazine.org/economy/2011/11/18/time-banking-an-idea-whose-time-has-come>

Optimalizácia topológií rozsiahlych sietí s efektívnou distribúciou komunikačných a streamingových služieb

¹Michal KUCAN, ²František JAKAB, ²Dávid CYMBALÁK,
²Roman VÁPENÍK

¹Katedra počítačov a informatiky, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

²Pracovisko centrálného manažmentu Národnej teleprezentačnej infraštruktúry, Centrum vedecko-technických informácií SR, Slovenská republika

michal.kucan@student.tuke.sk, frantisek.jakab@cvtisr.sk, david.cymbalak@cvtisr.sk,
roman.vapenik@cvtisr.sk

Abstrakt – Tento článok sa zaoberá implementáciou a simuláciou sieťových topológií, zameraných na poskytovanie komunikačných a streamingových služieb a odporúčaniami pre ich optimalizáciu. V rámci tohto článku sú implementované dve sieťové topológie na základe modelov definovaných spoločnosťou Cisco, pričom v jednej sú hovory spracovávané distribuovaným spôsobom a v druhej sú hovory spracovávané centralizovaným spôsobom. Tieto sieťové topológie sú následne simulované prostredníctvom postupne zvyšujúcej sa záťaže, pričom sa sleduje správanie sa ich kvalitatívnych parametrov, akými sú odozva, jitter a strata paketov. Na základe výsledkov simulácie sú na záver navrhnuté zlepšenia pre optimalizáciu poskytovania komunikačných a streamingových služieb v týchto topológiách.

Kľúčové slová – optimalizácia topológií, RTP, SIP, streaming, VoIP

I. ÚVOD

Protokol Voice over IP je v súčasnosti jednou z najdominantnejších technológií v komunikačnom svete. Predstavuje najjednoduchší spôsob, akým je možné telefonovať prostredníctvom internetu [1], a to pomocou posielania paketov cez paketovo prepínané siete. V globálnom kontexte, množstvo spoločností už dokázalo implementovať túto technológiu [1] a používať ju ako vhodnú náhradu k tradičnému analógovému telefónnemu systému.

VoIP technológia prináša množstvo výhod [2], či už sa jedná o úsporu nákladov, poskytovanie nových mediálnych služieb, prenositeľnosť alebo integráciu s inými aplikáciami. Napriek týmto výhodám, VoIP technológia trpí mnohými prekážkami [1], ako napríklad zložitosť jej architektúry alebo problémy s kvalitou služieb.

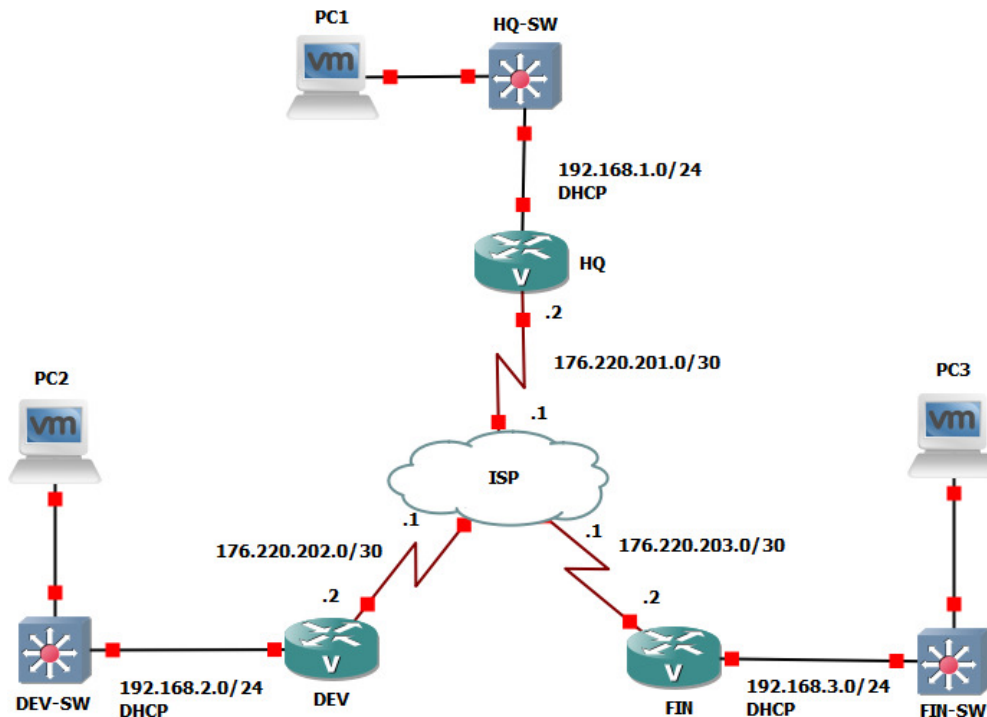
Hlavným dôvodom je, že hlasová prevádzka je charakterizovaná prísnyimi požiadavkami na kvalitu služieb. Kvôli povahe sieťovej technológie IP sú dátové pakety odosielané cez tento typ sietí vystavené určitým problémom s prenosom [3], ako je zvýšená odozva paketov, jitter, respektíve kolísanie odozvy paketov alebo strata paketov. Tieto problémy dokážu negatívne ovplyvniť konečný dojem z komunikácie prostredníctvom VoIP technológie. Aby mohli byť tieto problémy minimalizované, tak je žiadúce zaviesť mechanizmy, ktoré by riešili tieto problémy, a to v čo najväčšej možnej miere. Tomu sa venujú ďalšie kapitoly tohto článku.

II. IMPLEMENTÁCIA SIEŤOVÝCH TOPOLÓGIÍ

V rámci publikovaných kníh a dokumentácií, spoločnosť Cisco vo svojej príručke s názvom “Cisco IP Telephony Solution Reference Network Design“ jednotlivou opisuje modely, na základe ktorých môžu byť vytvorené VoIP sieťové topológie. Príručka nielen opisuje tieto modely z hľadiska princípu ich fungovania, ale poskytuje aj rôzne odporúčania a obmedzenia pre ich realizáciu. Pre účely tohto článku boli na základe týchto modelov navrhnuté a implementované sieťové topológie.

A. Sieťová topológia s distribuovaným spracovaním hovorov

Model siete s distribuovaným spracovaním hovorov podľa opisu v príručke od spoločnosti Cisco [4] pozostáva z viacerých nezávislých lokalít, pričom každá z týchto lokalít má svojho vlastného agenta pripojeného k WAN sieti, určeného pre spracovanie hovorov. Každý agent prenáša hlasovú prevádzku medzi distribuovanými lokalitami. Podľa tohto modelu bola navrhnutá a implementovaná sieťová topológia v simulačnom softvéri GNS3 [5], ktorú je možné vidieť na obrázku nižšie:



Obr. 1 Sieťová topológia s distribuovaným spracovaním hovorov [5]

Túto sieťovú topológiu tvoria štyri smerovače typu Cisco 3745. Smerovače pomenované ako HQ, DEV a FIN reprezentujú jednotlivé lokality [5] tejto sieťovej topológie. Smerovač s názvom ISP zase reprezentuje poskytovateľa internetových služieb [5] a umožňuje komunikáciu medzi všetkými lokalitami.

Okrem služieb smerovania a poskytovania IP adries koncovým bodom, každý smerovač v lokalitách HQ, DEV a FIN poskytuje aj VoIP služby v svojej lokalite prostredníctvom platformy Cisco Unified Communications Manager Express [5], čím je dodržaný distribuovaný princíp spracovania hovorov. Pre poskytovanie VoIP služieb [5], v rámci tejto sieťovej topológie bol použitý SIP ako signalizačný protokol.

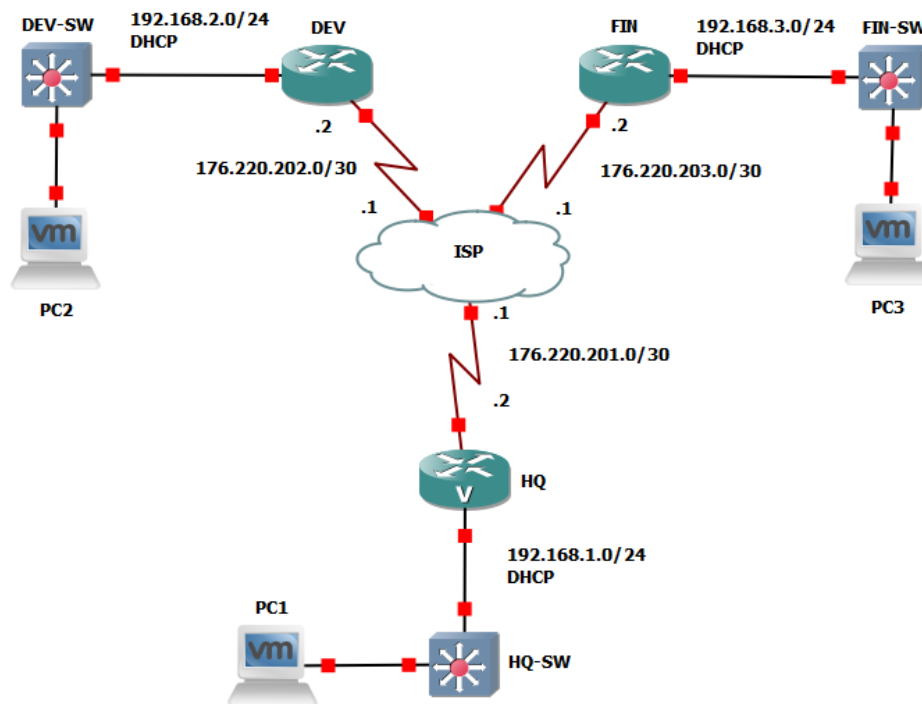
Smerovač ISP umožňuje komunikáciu medzi všetkými lokalitami v rámci tejto sieťovej topológie. Smerovač je prepojený s každou lokalitou prostredníctvom sériových liniek [5], pričom na jej koncoch sú nakonfigurované IP adresy z iných než privátnych rozsahov IP adries. Smerovanie medzi jednotlivými lokalitami je na tomto smerovači zabezpečené prostredníctvom BGP smerovacieho protokolu [5].

V každej lokalite je na smerovač pripojený Cisco IOSvL2 prepínač, ktorý prepája koncové body so smerovačom [5], aby mohli komunikovať s ostatnými lokalitami. Koncové body sú v tejto topológii reprezentované virtuálnymi počítačmi [5], na ktorých je nainštalovaný operačný systém Windows. V rámci Windowsu je nainštalovaný softvér MicroSIP [5], čo je softphone softvér umožňujúci uskutočňovať VoIP hovory. Dohromady sa v tejto sieťovej topológii nachádzajú tri koncové body [5], respektíve traja VoIP klienti.

B. Sieťová topológia s centralizovaným spracovaním hovorov

Na rozdiel od modelu sieťovej topológie s distribuovaným spracovaním hovorov, tento model pozostáva z jedinej lokality obsahujúcej agenta spracovania hovorov [4], ktorý poskytuje služby pre všetky lokality v sieťovej topológii a využíva WAN sieť na prenos hlasovej prevádzky medzi lokalitami. WAN sieť taktiež prenáša signalizáciu riadenia hovorov medzi centrálnou lokalitou

a vzdialenými lokalitami [4]. Podľa tohto modelu bola navrhnutá a implementovaná sieťová topológia v simulačnom softvéri GNS3 [5], ktorú je možné vidieť na obrázku nižšie:



Obr. 2 Sieťová topológia s centralizovaným spracovaním hovorov [5]

Táto sieťová topológia je prakticky totožná s topológiou siete s distribuovaným spracovaním hovorov z hľadiska použitých sieťových prvkov a prepojení medzi nimi [5]. Aj v tomto prípade sa tu nachádzajú štyri smerovače typu Cisco 3745, ktoré majú rovnaké pomenovanie ako aj v sieťovej topológii s distribuovaným spracovaním hovorov [5] a reprezentujú jednotlivé lokality, respektíve poskytovateľa internetových služieb. Rovnako sa v každej lokalite nachádza Cisco IOSvL2 prepínač, ktorý takisto prepája koncové body so smerovačom [5], aby mohli komunikovať s ostatnými lokalitami. Koncové body sú opäť reprezentované virtuálnymi počítačmi, na ktorých je nainštalovaný operačný systém Windows spoločne so softvérom MicroSIP a takisto sa v každej lokalite nachádza len jeden koncový bod [5].

Na rozdiel od sieťovej topológie s distribuovaným spracovaním hovorov [5], v tejto sieťovej topológii je možné objaviť jeden podstatný rozdiel, a to že VoIP služby poskytuje iba smerovač v lokalite HQ. Je teda zodpovedný za poskytovanie VoIP služieb pre celú sieť a VoIP klienti, nachádzajúci sa v ostatných lokalitách, sa musia registrovať v systéme tohto smerovača, bežiacom na platforme Cisco Unified Communications Manager Express, aby mohli uskutočňovať VoIP komunikáciu [5].

III. SIMULÁCIA SIEŤOVÝCH TOPOLÓGIÍ

Pri oboch sieťových topológiách bola vykonaná simulácia ich sieťovej prevádzky, či už bez záťaže alebo s postupne sa zvyšujúcou záťažou. Simulácia ako taká bola vykonávaná tak, že bol vždy uskutočnený testovací hovor medzi koncovými bodmi PC1 v lokalite HQ a PC2 v lokalite DEV, pričom záťaž bola simulovaná prostredníctvom nástroja iPerf, ktorý slúži na meranie výkonnosti siete a dokáže generovať prúdy dátových paketov medzi dvoma bodmi v sieti [5]. Pre účely simulácie záťaže bol určený koncový bod PC1 ako klient, ktorý odosielať dátové pakety a koncový bod PC2 zohral úlohu servera, ktorý tieto dátové pakety prijímal [5].

V každej sieťovej topológii bolo vykonaných dohromady 5 testovacích hovorov [5], počas ktorých bola vyvíjaná záťaž nasledovne:

- prvý testovací hovor - žiadna záťaž,
- druhý testovací hovor - záťaž v podobe 1 ďalšieho hovoru,
- tretí testovací hovor - záťaž v podobe 5 ďalších hovorov,
- štvrtý testovací hovor - záťaž v podobe 10 ďalších hovorov,
- piaty testovací hovor - záťaž v podobe 20 ďalších hovorov.

Počas týchto testovacích hovorov boli pomocou softvéru Wireshark zachytávané prichádzajúce RTP streamy [5], u ktorých boli sledované kvalitatívne parametre. Menovite sa jednalo o odozvu paketov, jitter a stratu paketov. Namerané hodnoty týchto parametrov počas simulácie oboch sieťových topológií je možné vidieť v tabuľkách nižšie:

Sieťová topológia s distribuovaným spracovaním hovorov					
Typ záťaže	Očakáv. paketov	Strat. paketov	Max. odozva	Priem. jitter	Max. jitter
Bez záťaže	1493	0	32,49 ms	1,71 ms	3,58 ms
1 hovor	3014	0	41,88 ms	2,59 ms	6,18 ms
5 hovorov	2368	450	103,55 ms	4,77 ms	18,51 ms
10 hovorov	2811	764	185,91 ms	4,01 ms	41,46 ms
20 hovorov	2878	1339	658,58 ms	4,19 ms	64,34 ms

Tabuľka 1

Hodnoty kvalitatívnych parametrov sieťovej topológie s distribuovaným spracovaním hovorov

Sieťová topológia s centralizovaným spracovaním hovorov					
Typ záťaže	Očakáv. paketov	Strat. paketov	Max. odozva	Priem. jitter	Max. jitter
Bez záťaže	1614	0	33,69 ms	1,69 ms	4,04 ms
1 hovor	2557	0	40,49 ms	2,74 ms	5,95 ms
5 hovorov	2668	453	92,3 ms	4,2 ms	18,52 ms
10 hovorov	2793	767	192,9 ms	4,21 ms	41,31 ms
20 hovorov	2681	1254	568,79 ms	8,12 ms	140,96 ms

Tabuľka 2

Hodnoty kvalitatívnych parametrov sieťovej topológie s centralizovaným spracovaním hovorov

Pri pohľade na obidve tabuľky nameraných hodnôt kvalitatívnych parametrov je možné všimnúť si, že s narastajúcou záťažou bolo možné pozorovať zhoršené správanie sa oboch sieťových topológií počas testovacích hovorov, čo viedlo k ich zhoršenej kvalite a konečného používateľského zážitku. Rovnako je možné konštatovať, že tieto sieťové topológie sa pri rôznych úrovniach záťaže správali prakticky rovnako, s výnimkou testovacieho hovoru [5], počas ktorého bola simulovaná záťaž ďalšími piatimi súbežnými hovormi. V tomto prípade sa hodnoty jitteru odlišovali v neprospech sieťovej topológie s centralizovaným spracovaním hovorov, kde boli namerané hodnoty jitteru dvojnásobné v porovnaní so sieťovou topológiou s distribuovaným spracovaním hovorov [5].

IV. ODPORÚČANIA PRE OPTIMALIZÁCIU SIEŤOVÝCH TOPOLÓGIÍ

Simulácia oboch sieťových topológií, či už bez záťaže alebo so záťažou, ukázala ich slabiny [5], ktoré ovplyvňujú VoIP komunikáciu nielen z hľadiska jej kvality, ale aj z hľadiska samotného používateľského zážitku. Na základe týchto skutočností je možné konštatovať [5], že existuje priestor na zlepšenie výkonnosti oboch sieťových topológií.

Pre zlepšenie výkonnosti sieťových topológií počas VoIP komunikácie je možné vykonať konfiguráciu QoS mechanizmu [5], ktorý by bol sústredený na riadenie sieťovej prevádzky a tým pádom zabezpečoval čo najvyšší výkon počas prebiehajúcej VoIP komunikácie. Medzi jeden zo spôsobov konfigurácie QoS mechanizmu je funkcia "Modular QoS CLI", ktorá je dostupná v rámci sieťových prvkov od spoločnosti Cisco a dovoľuje označovať a klasifikovať pakety v sieti na základe ich vlastností [5].

Ďalšou z možností pre zlepšenie výkonnosti sieťových topológií je zmena preferovaného kodeku, ktorý by sa používal počas prebiehajúcej VoIP komunikácie, a to najmä vtedy, ak sa v sieti pracuje s obmedzenou šírkou pásma [5]. V rámci oboch sieťových topológií bol ako preferovaný kodek použitý G.711 [5], ktorý sa vyznačuje bitovou rýchlosťou 64 kilobitov za sekundu. Ako alternatívu je možné použiť kodek G.729, ktorý sa voči kodeku G.711 vyznačuje nízkou bitovou rýchlosťou, a to len 8 kilobitov za sekundu. Tú dosahuje prostredníctvom stratovej kompresie dát, na rozdiel od kodeku G.711, ktorý nijakým spôsobom nekomprimuje dáta [5].

Rovnako môže byť výkonnosť sieťových topológií ovplyvnená nielen nesprávnou konfiguráciou parametrov alebo chýbajúcou konfiguráciou funkcií a mechanizmov, ale aj použitými sieťovými prvkami [5]. V kontexte tohto článku sa zistilo, že použité prepínače Cisco IOSvL2 sú obmedzované v rýchlosti presmerovania sieťovej prevádzky a s najväčšou pravdepodobnosťou táto skutočnosť ovplyvnila simuláciu oboch sieťových topológií.

V. ZÁVER

V tomto článku bola znázornená implementácia a simulácia sieťových topológií, ktoré sú zamerané na poskytovanie komunikačných a streamingových služieb. Pre tieto účely boli v simulačnom softvéri GNS3 implementované dve sieťové topológie, pričom v jednej boli hovory spracovávané distribuovaným spôsobom a v druhej boli hovory spracovávané centralizovaným spôsobom. Tieto sieťové topológie boli následne simulované na základe postupne zvyšujúcej sa záťaže pomocou nástroja iPerf, pričom sa pomocou softvéru Wireshark sledovalo správanie sa ich kvalitatívnych parametrov, akými su odozva paketov, jitter a strata paketov. Výsledky tejto simulácie ukázali, že obidve sieťové topológie vykazujú slabiny, ktoré negatívne ovplyvňujú VoIP komunikáciu, a teda aj konečný používateľský zážitok počas takejto komunikácie. Na základe týchto poznatkov boli navrhnuté odporúčania pre zvýšenie výkonnosti týchto sieťových topológií a minimalizáciu ich problémov.

VI. POĎAKOVANIE

Táto publikácia vznikla vďaka podpore v rámci Operačného programu Integrovaná infraštruktúra pre projekt: Výskum v sieti SANET a možnosti jej ďalšieho využitia a rozvoja. ITMS kód: 313011W988, spolufinancovaný zo zdrojov Európskeho fondu regionálneho rozvoja.

LITERATÚRA

- [1] O. Hersent, J.-P. Petit, and D. Gurle, *Beyond VoIP protocols: understanding voice technology and networking techniques for IP telephony*. Hoboken, NJ: John Wiley, 2005.
- [2] R. Vápeník, “Zabezpečený webovo-orientovaný videokonferenčný systém,” Master’s thesis, Technická univerzita v Košiciach, 2014.
- [3] R. Vápeník, “Metódy vyhodnocovania parametrov streamingu na báze sledovacích algoritmov v multikamerových systémoch,” Ph.D. dissertation, Technická univerzita v Košiciach, 2020.
- [4] *Cisco IP Telephony Solution Reference Network Design*, Cisco Systems, 170 West Tasman Drive, San Jose, CA 95134-1706, 1 2005.
- [5] M. Kucan, “Optimalizácia topológií rozsiahlych sietí s efektívnou distribúciou komunikačných a streamingových služieb,” Master’s thesis, Technická univerzita v Košiciach, 2022.

Open-source platformy pre elektroenergetické účely

Gabriel BARAN, Dušan MEDVEĎ

Katedra elektroenergetiky, Fakulta elektrotechniky a informatiky,
Technická univerzita v Košiciach, Slovenská republika

gabriel.baran@student.tuke.sk, dusan.medved@tuke.sk

Abstrakt — Tento príspevok sa zaoberá návrhom inteligentnej inštalácie využitím open source platformy, v našom prípade Arduino MEGA 2560, ktorú sme využili pri návrhu inteligentného domu. Tento príspevok je rozdelený do niekoľkých častí zameraných na samotné riadenie Arduino senzorov, ktoré boli pre účely príspevku využité. V tomto článku je uvedený popis využitia jednotlivých senzorov, ich princíp činnosti, samotné meranie a vyhodnotenie výsledkov z merania riadiacej dosky Arduino MEGA 2560.

Kľúčové slová — riadiaca jednotka, Arduino MEGA 2560, inteligentný dom, DHT11

I. ÚVOD

Využitie open-source platformiem v dnešnej dobe je zaujímavou myšlienkou najmä z hľadiska ekonomického, pretože open source programy sú voľne stiahnuteľné vo webovom prehliadači a môžeme ich vlastne využívať zadarmo. Našou voľbou bola vývojová doska Arduino MEGA 2560, ktorá predstavuje voľne dostupnú a šíriteľnú platformu. Pomocou tejto dosky sme navrhli prototyp reprezentujúci inteligentný dom.

II. OPEN SOURCE PLATFORMY

Open source je vo všeobecnosti akákoľvek informácia dostupná verejnosti, za podmienky, že možnosť jej slobodného šírenia zostane zachovaná. V tomto prípade nie je účtovaný poplatok za obdržanie, open source znamená slobodu prístupu. Open source je filozofia hnutia založeného Ericom Raymondom.

Open source softvér (angl. open-source software (OSS)) je počítačový softvér s otvoreným zdrojovým kódom, čo znamená, že používatelia musia mať právo voľne ho používať, modifikovať a šíriť, ako aj právo predávať ako produkt. Tieto podmienky spĺňa softvér uvoľnený pod licenciami GNU General Public License (GPL), public domain, GPL, BSD a podobne.

Používanie OSS softvéru okrem toho, že je mimoriadne efektívny z hľadiska nákladov, poskytuje oproti plateným alternatívam niektoré významné výhody:

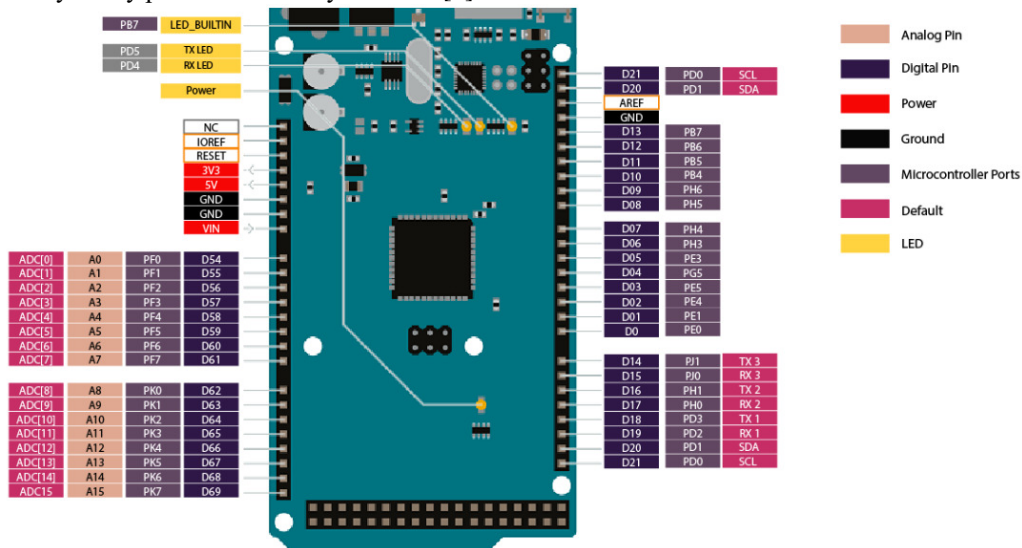
- *Bezpečnosť* – vzhľadom na skutočnosť, že kód programu je verejne dostupný, býva často skúmaný odborníkmi po celom svete, bezpečnostné chyby teda bývajú rýchlejšie nájdené a aj opravené.
- *Kontrola* – vďaka otvorenému kódu môže používateľ presne zistiť, čo program robí. Má teda plnú kontrolu nad procesom.
- *Stabilita* – open source programy sa vyznačujú dlhodobou podporou, vývojom a aktualizáciami.
- Platformy predstavujú aj značné nedostatky a preto je nutná *neustála manuálna údržba, sledovanie updatov a ich nahrávanie*. Tieto činnosti si vyžadujú zásah programátora, čím sa prevádzka www stránky alebo e-shopu predraží.
- *Pluginy alebo zmeny vyžadujú programátora*. Rovnako, ako v predchádzajúcom bode, pokiaľ nepatríte medzi programátorov, nebudete si zrejme vedieť naimplementovať ani pluginy (napr. rôzne SEO balíčky alebo Broken Link Checker a pod.), ani čokoľvek zmeniť. Výmena hlavičky alebo slajdov v banneri vám opäť zasiahne do rozpočtu. [1]

III. ARDUINO MEGA 2560

A. Arduino MEGA 2560

Arduino MEGA 2560 je doska mikrokontroléra založená na Atmega2560. Dodáva sa s väčším pamäťovým priestorom a I/O pinmi v porovnaní s inými doskami dostupnými na trhu. Na doske je zabudovaných 54 digitálnych I/O pinov a 16 analógových pinov, vďaka ktorým je toto zariadenie jedinečné a vyniká od ostatných. Z 54 digitálnych I/O pinov sa 15 používa na PWM (pulzne šírková modulácia). Na doske je pridaný kryštálový oscilátor s frekvenciou 16 MHz. Táto doska sa dodáva pripojením na USB, ktorý sa používa na pripojenie a prenos kódu z počítača do dosky. DC napájací konektor je spojený s doskou, ktorá sa používa na napájanie mikrokontroléra. Niektorým verziám dosky Arduino chýba táto funkcia, napríklad Arduino Pro Mini sa nedodáva s DC napájacím konektorom. ICSP hlavička je pozoruhodný doplnok k Arduino Mega, ktorý sa používa na programovanie Arduina a nahrávanie kódu z počítača. Táto doska je dodávaná s dvoma regulátormi napätia, t.j. 5 V a 3,3 V, ktoré poskytujú flexibilitu pri regulácii napätia podľa požiadaviek v porovnaní s Arduino Pro Mini, ktorý sa dodáva iba s jedným regulátorom napätia.

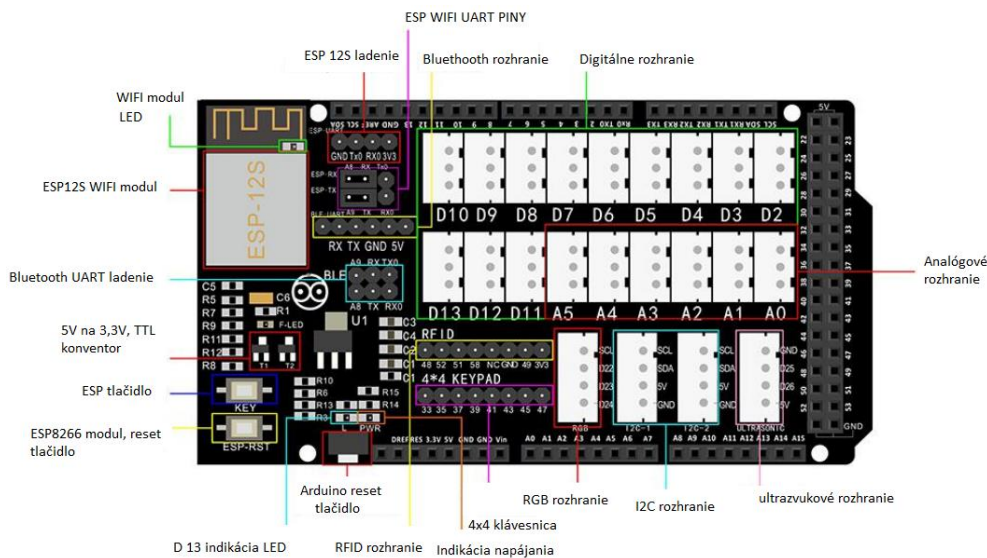
Medzi Arduino Uno a Arduino Mega nie je veľký rozdiel, okrem toho, že neskôr prichádza s väčším pamäťovým priestorom, väčšou veľkosťou a viacerými I/O pinmi. Softvér Arduino s názvom Arduino IDE sa používa na programovanie dosky, čo je bežný softvér používaný pre všetky dosky patriace do rodiny Arduino [2].



Obr. 1 Výstupné analógové a digitálne piny Arduino MEGA2560

B. Arduino MEGA IoT-shield

Arduino MEGA-IoT shield je rozširovacia Arduino doska, ktorá integruje WiFi čip ESP8266 a poskytuje lacné WiFi riešenie pre akékoľvek projekty Arduino MEGA2560. Arduino MEGA-IoT Shield je plug and play doska (vloží a hraje), je kompatibilná s doskou Arduino MEGA2560, kde sa jednoducho položí doska na Arduino Mega 2560. Taktiež je potrebné vybrať piny pre sériovú komunikáciu pomocou mini prípojok a hardvérové pripojenie je dokončené. ESP8266 je modul UART WiFi s ultranízkou spotrebou. Arduino MEGA-IoT Shield je založený na ESP-12s s čipom ESP8266 WiFi. Tento shield je určený na jednoduché pripojenie Arduina k internetu (IoT). Poskytuje cenovo dostupné riešenie WiFi s akýmkoľvek projektmi Arduino MEGA2560. Kvôli internetu vecí (Internet of Thing) by malo byť všetko pripojené k internetu, kľúčovým prvkom je WiFi, ktoré poskytuje bezdrôtové pripojenie k internetu. Pre vývoj programu, existuje veľa knižníc, ktoré môžeme využiť. ESP12S sa dodáva s predflashovaným firmvérom AT-command, takže ho možno ovládať cez akýkoľvek UART/Serial, ale tiež sa spína a poskytuje príkazový prístup k všetkým I/O ESP8266. Samotný shield sa neobmedzuje na AT príkaz, kde môžeme priamo naprogramovať modul ESP12S. V prípade, že chceme pridať AT príkaz pomocou vlastných príkazov alebo flash vlastného firmvéru na ESP8266, tak táto funkcia sa najmä hodí s využitím pinových voličov UART [3].



Obr. 2 Arduino MEGA IoT rozširovacia doska

IV. INICIALIZÁCIA ARDUINO DOSKY S WIFI MODULOM

Pri prepojení komunikačného protokolu http s našou riadiacou jednotkou domu, čo je vlastne Arduino2560 a rozširovacia jednotka, resp. doska Arduino IoT Shield, bolo dôležité v programe zahrnúť knižnicu WiFiEsp.h. Pre inicializáciu a prepojenie Arduina s webovým prehliadačom bolo potrebné získať údaje ohľadom nášho routera, kde sme museli získať názov SSID a heslo. Informácie potrebné na inicializáciu Arduina s webovým prehliadačom sme mohli nájsť na zadnej strane routera, išlo o štítkové údaje. Čo sa týka, zapojenia samostatných pinov riadiacej jednotky, museli sme riešiť situáciu, kde pri zapojení a inicializovaní pinov A8 – RX a A9 – TX (prepojenie na riadiacej doske Arduina) bol čas na prepojenie s internetovým prehliadačom nedostatočný a prepojenie medzi http protokolom a riadiacou doskou Arduina bol nedostatočný a sériový monitor nám vypísal: TIMEOUT, čo predstavovalo, že vysielačie a prijímacie piny A8 a A9 nie sú stabilné a dostatočne rýchle, predstavovali analógové piny, ktoré sme museli nahradiť digitálnymi, ich výstup predstavovali piny D10 a D11, takže nasledujúce prepojenie pinov bolo RX-D10, TX-D11.

Zo seriového monitora môžeme načítať IP adresu nášho mikrokontroléra Arduina. Prostredníctvom zadania príslušného linku vieme zadaný text „Hello world“ zobrazit' v našom prehliadači. Dané riešenie potom značí vstupné napätie kolíka A0 prostredníctvom funkcie analogRead(0) inline94. analogRead() dokáže prečítať vstupné napätie 0÷5 V a vrátiť hodnotu medzi 0 až 1023, ktorá predstavuje napätie. Pripojili sme A0 na 5 V, takže nad analógovým vstupom A0 ukazuje 1023.

Aby naše senzory boli prepojené s internetom, musíme zabezpečiť komunikáciu a rozpoznanie medzi riadiacimi senzormi a komunikačným protokolom. Medzi základné typy protokolov patria IoT, HTTP, MQTT. Pre účely tohto príspevku sme pracovali so štandardným http protokolom. Protokol nám tiež, dokáže zobrazit' vstupné napätie na pine A0, ktoré je detegované funkciami Arduino analogRead() a zároveň nám predstavuje naše napájacie napätie 5 V, ale vzhľadom na skutočnosť, že analógové piny pracujú pomalšie, museli sme ich nahradiť za digitálne (10,11).

V. SNÍMAČ TEPLoty A VLHKOSTI DHT11

DHT11 je základný, veľmi lacný digitálny snímač teploty a vlhkosti. Na meranie okolitého vzduchu využíva kapacitný snímač vlhkosti a termistor, a na dátový pin vysiela digitálny signál (nie sú potrebné žiadne analógové vstupné piny). Jeho použitie je pomerne jednoduché, ale vyžaduje si starostlivé načasovanie na získanie údajov, lebo jeho skutočnou nevýhodou je, že z neho môžeme získať nové údaje iba raz za 2 sekundy, pričom údaje snímača môžu byť navzorkované s diskretizáciou až 2 sekundy. V porovnaní s DHT22, tento senzor je menej presný, menej citlivý a pracuje v menšom rozsahu teploty/vlhkosti, ale je menší a lacnejší.

A. Relatívna vlhkosť

Tento senzor DHT11 meria relatívnu vlhkosť. Relatívna vlhkosť je množstvo vodnej pary vo vzduchu v porovnaní s bodom nasýtenia vodnej pary vo vzduchu. V bode nasýtenia vodná para začne kondenzovať a hromadiť sa na povrchoch a vytvárať kondenzát (rosu). Bod nasýtenia sa mení s teplotou vzduchu. Studený vzduch pojme menej vodnej pary, kým sa nasýti, a horúci

vzduch pojme viac vodnej pary, kým sa nasýti [4].

$$R_H = \frac{\rho_w}{\rho_s} \cdot 100\% \quad (1)$$

kde: R_H relatívna vlhkosť
 ρ_w hustota vodnej pary
 ρ_s hustota vodnej pary pri nasýtení

VI. VYUŽITIE MOBILNEJ APLIKÁCIE

Pre ovládanie danej inštalácie s využitím Arduino Mega, bolo potrebné stiahnutie mobilnej aplikácie (na platforme Android), ktorá v sebe zahŕňala známy protokol UDP, ktorý bol využitý pre prenos a riadenie signálu (PWM). Taktiež bolo potrebné stiahnutie aplikácie „WiFi UDP Robot car“. Pre ovládanie aplikácie boli potrebné ešte dve knižnice servo.h a keypad.h, vzhľadom na to, že servo-motor bolo potrebné riadiť cez aplikáciu a taktiež hodnoty natočenia servo-motora je možné nastavovať prostredníctvom klávesnice Keypad.

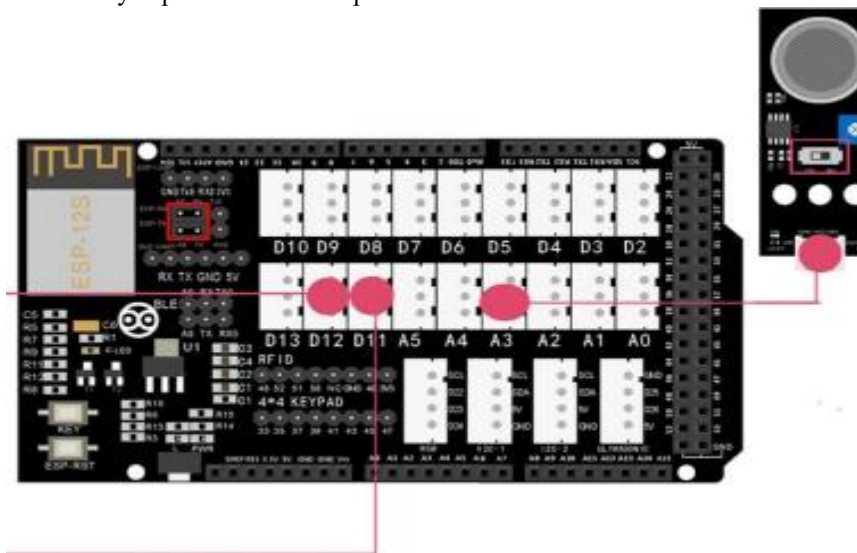
Pri nastavovaní aplikácie bolo potrebné v prvom kroku stiahnuť aplikáciu WiFi UDP Robot car, pomocou ktorej sme nastavovali polohy, t.j. kde a o koľko stupňov sa má servo-motor natočiť. Prvotná inicializácia aplikácie spočívala v tom, že sme museli mať prístup na internet, taktiež sme sa museli pripojiť k WiFi sieti a postupne sme urobili prvotné nastavenia aplikácie a prepojili ju s Arduino. Pri inicializácii kódu je možné vidieť, že sériový port COM3 vypíše IP adresu a port. Tieto 2 údaje sú dôležité a je ich potrebné zadať pri nastavení a prepojení aplikácie s Arduino. Pri riadení domu, pomocou tejto aplikácie, je dôležité mať permanentne zabezpečené sieťové pripojenie, aby sme dokázali zabezpečiť komunikáciu dosky s aplikáciou.

VII. PRIPOJENIE JEDNOTLIVÝCH SENZOROV

Pre účely tohto príspevku bude v tejto časti uvedený princíp zapojenia plynového senzora do inteligentnej inštalácie. Ďalšie zapojenia senzorov si môže čitateľ nájsť v referencii [6], kde je uvedená aj celkové zapojenie a testovanie príslušných snímačov.

A. Plynový senzor

Jedným z dôležitých senzorov z Arduino ponuky, je plynový senzor, ktorý slúži na detegovanie plynu a taktiež aj dymu v miestnosti. Tento senzor je umiestnený na pravej bočnej stene a pomocou neho sme vyhodnocovali, či sa v miestnosti nachádza alebo nenachádza plyn alebo dym. Výstupy boli realizované cez vzdialený webový prehliadač. Vzhľadom na to, že sme Arduino inicializovali a prepojili s WiFi modulom, bolo možné využiť toto riešenie prostredníctvom výstupov vo webovom prehliadači.



Obr. 3 Zapojenie plynového senzora

Na obrázku 3 môžeme vidieť schému zapojenia plynového senzora a spolu s ním aj zapojenie červenej a zelenej LED diódy, ktoré budú slúžiť na signalizáciu a vyhodnocovanie stavu v miestnosti resp. v dome, kde jednotlivé farby znamenajú:

- Červená LED – signalizuje, že v dome bolo detegované určité množstvo plynu resp.

dymu, LED sa rozsvieti.

- Zelená LED – signalizuje, bezpečný status, v prípade, že nie je detegovaný žiadny plyn alebo dym, tak sa LED rozsvieti na zeleno.

Plynový senzor fungoval v 2 polohách:

- D – deaktivovaný, nulová detekcia dymu
- A – aktivovaný, aktivovaný plyn v miestnosti

Samotnú realizáciu prepojenia senzora spolu s LED diódami k riadiacej jednotke domu Arduino sme nastavili pomocou jednoduchšej podmienky „if, else“, kde sme zadefinovali jednotlivé výstupy senzora a LED diód. Pri detekcii plynu, resp. dymu nachádzajúceho sa v miestnosti boli nasledovné stavy: Ak plynový senzor bol v stave 1 (poloha D), bolo všetko v poriadku, žiadna detekcia plynu v dome, zelená LED sa rozsvieti a výstup na prehliadači nám hlási bezpečnostný stav. V opačnom prípade, keď sa plynový senzor aktivoval (poloha A), došlo k rozsvieteniu červenej LED diódy a výstup nám hlásil detekciu nežiadúcich vplyvov v miestnosti.

```
void loop()
{
    gasStatus=digitalRead(gasSensor);
    if (gasStatus==1) {
        digitalWrite(redLED,LOW);
        digitalWrite(greenLED,HIGH);
        gasStr="<font color=GREEN><b>bezpecne</b></font>";
    } else
    {
        digitalWrite(redLED,HIGH);
        digitalWrite(greenLED,LOW);
        gasStr="<font color=RED><b>Detekcia plynu!</b></font>";
    }
}
```

Obr. 4 Funkcia pre nastavenie plynového senzoru – zdrojový kód pre Arduino

VIII. MERANIE SPOTREBY ARDUINO MEGA 2560

Pri meraní a vyhodnocovaní spotreby riadiacej jednotky Arduino Mega2560 bolo potrebné vykonať na doske Mega 2560 určité úpravy, aby sme dosiahli, čo najnižšiu spotrebu.

V prvom kroku, k dosiahnutiu nižšej spotreby dosky Arduino, bolo potrebné odstránenie, resp. odpojenie z obvodu LED indikácie napájania. S indikátorom je totiž spotreba vyššia, čo indikuje prechádzajúci prúd cca 46 mA a našim cieľom bolo zníženie spotreby. Preto bolo odpojenie LED nutnosťou. Ďalším krokom bolo pre úpravu spotreby dôležité odpojenie regulátoru napätia, tiež mal svoju spotrebu, ktorú sme v našom prípade zanedbali odpojením.

A. Režim spánku (Power Down)

Ako prvé sme vyhodnocovali spotrebu zariadenia v režime spánku (najmenšia spotreba). Mikroprocesor sme teda uviedli do najhlbšieho spánku s najnižšou spotrebou. Pri meraní spotreby sme využili knižnicu Low-Power a uspávali sme mikroprocesor na 8 sekúnd, aby náš merací prístroj stihol stabilizovať nameranú hodnotu. Pri metóde uspávania mikroprocesora mali vplyv na spotrebu aj dva parametre – ADC (Analogue to Digital Converter) – analógovo-digitálny prevodník a BOD (Brown Out Detector) – detekcia poklesu napätia. Meranie spotreby sme vykonali s vypnutými aj so zapnutými parametrami ADC a BOD.

Tabuľka 1
 Meranie spotreby – spánkový režim

Stav			Spotreba
	ADC	BOD	[μ A]
PWRdown	✓	✓	116,9
PWRdown	✓	✗	110,2
PWRdown	✗	✓	23,3
PWRdown	✗	✗	4,3

Z tabuľky 1 je zrejmé, že režim Power Down, s vypnutými ADC aj BOD, má výsledok spotreby 4,3 μ A. Výsledok by sa dal ešte mierne vylepšiť, keby sme namiesto WDT (časovač WatchDog) použili zobudzanie externým obvodom, ale pridávať len kvôli tomu externý časovač môže v konečnom dôsledku spotrebu zvýšiť, samozrejme pokiaľ tento obvod nie je na činnosť zariadenia nutný [5].

B. Režim za behu

Pre nižšiu spotrebu „za behu“, má mikroprocesor možnosť vypínania jednotlivých súčastí

mikroprocesora. Pri našom type procesora sa jedná o: komunikačné modely SPI, IIC (TWI), UART. Moduly časovačov: Timer0, Timer1, Timer2, modul ADC. Funkcie jednotlivých zariadení sú súčasťou knižnice AVR Libc. Taktiež sme vypínali časovač Timer0, ktorý Arduino používa na čakacie slučky (delay()). Z tohto dôvodu sme museli použiť čakaciu slučku z AVR LibC, tá však podporuje max 6,5 sekundy. V tomto prípade, LED dióda bliká a pri každom zabliknutí sa vypína/zapína určité zariadenie a taktiež samostatne vypíname len ADC. V programe funkcie powerOn() a powerOff() jednoducho zapínajú a vypínajú moduly v tomto poradí: 1.SPI, 2. IIC (TWI), 3. UART, 4. Časovač 0, 5. Časovač 1, 6. Časovač 2 [5].

Tabuľka 2
Meranie spotreby – režim za behu

Stav	ADC	Ostatné On	Ostatné Off
		[mA]	[mA]
LED On	✓	4,9	4,5
LED On	✗	4,8	4,4
LED Off	✓	3,5	3,1
LED Off	✗	3,4	3,0

IX. ZÁVER

Cieľom tohto príspevku bolo predviesť výsledky návrhu domovej inštalácie s využitím prvkov inteligentného riadenia za použitia open-source platformy Arduino Mega 2560. Tento príspevok bol rozdelený do niekoľkých častí zameraných na samotné riadenie Arduino senzorov, ktoré boli pre účely príspevku využité. V tomto príspevku bol uvedený popis využitia niektorých senzorov, ich princíp činnosti, samotné meranie a vyhodnotenie výsledkov z merania riadiacej dosky Arduino MEGA 2560.

POĎAKOVANIE

Tento príspevok vznikol vďaka Agentúre na podporu výskumu a vývoja na základe zmluvy č. APVV-19-0576, ďalej Vedeckej grantovej agentúry Ministerstva školstva, vedy, výskumu a športu SR a SAV podporenej grantom VEGA 1/0757/21 a Kultúrnej a edukačnej grantovej agentúre MŠVVaŠ SR pod číslom KEGA 013TUKE-4/2021.

ZOZNAM POUŽITEJ LITERATÚRY

- [1] R. Taylor, "Open-source software and platforms", [online], [cit.: 1.5.2022], Chicago university library, dostupné na <https://opensource.com/resources/what-open-source>.
- [2] J. Pavlík, "Arduino Mega 2560", [online], [cit.: 4.5.2022] dostupné na <https://arduinoposlovensky.sk/hardware/arduino-mega2560>.
- [3] Arduino corporation, "arduinocc", [online], [cit.: 1.5.2022] dostupné na <https://www.arduino.cc/en/pmwiki.php?n=Main/ArduinoWiFiShield>.
- [4] arduino.cc, "projekthub", [online], [cit.: 4.5.2022] dostupné na <https://create.arduino.cc/projecthub/pibots555/how-to-connect-dht11-sensor-with-arduino-uno-f4d239>.
- [5] Diyloot, „Guide to reduce the Arduino Power Consumption“, [online], [cit.: 2.5.2022] dostupné na <https://diyloot.com/arduino-reduce-power-consumption/>.
- [6] Baran, G., „Open source platformy pre elektroenergetické účely“. Bakalárska práca. Technická univerzita v Košiciach, Fakulta elektrotechniky a informatiky, 2022.

Use of Recurrent Fuzzy Neural Networks in Control of Mobile Robots

Ján VAŠČÁK

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

jan.vascak@tuke.sk

Abstract – This paper deals with the use of recurrent fuzzy neural networks in the control of wheeled actuators of mobile robots. Starting from a kinematic description of the robot, a serial controller is proposed as a combination of a conventional fuzzy controller and a neural network. The design has been simulation tested and also compared against a conventional MLP network.

Keywords – fuzzy controller, mobile robot, navigation, recurrent fuzzy neural network

I. INTRODUCTION

The main topic of this paper is the navigation of a two-wheeled mobile robot using recurrent fuzzy neural networks (RFNNs) based on the work of [1] and [2]. Sequentially, we discuss the analytical description of a mobile robot model in a MATLAB environment and then the design of an RFNN controller to control the DC motor voltage of the robot's wheel.

The main motivation for linking fuzzy logic and neural network is their potential to complement each other in their properties as well as the ability of both means to approximate analytical functions. In this way, we have the possibility to ensure a faster convergence of the system behavior towards a prescribed state [3], [4].

II. DESCRIPTION OF RFNN STRUCTURE

The main drawback of classical fuzzy neural networks (FNNs) is the focus on specific static problems due to their forward structure. Therefore, for our type of problem, it is more appropriate to introduce a RFNN based on supervised learning (learning with a supervisor), and therefore it is more suitable for dynamic systems. This type of networks can handle time-varying input and it is also possible to temporarily store information in this way, resulting in a simpler network that has fewer links and nodes [5]. As it can be seen in Fig. 1, the RFNN is formed by a classical FNN, with the addition of a feedback loop implemented in the second layer of its structure. This modification ultimately provides memory elements and elimination of timing problems. The recurrent neuron contains the inner loop shown, which captures the dynamic response of the system [6].

The structure shown in Fig. 1 represents the fuzzy inference system used in our solution. Specifically, this is formed by n inputs in the first layer, m nodes for each input layer, p output nodes, and $m*n$ rule nodes. The RFNN system itself consists of four layers and $n + (n * m) + m + p$ nodes, where m is the number of rules. As mentioned layer 1 forms the input of the RFNN, which receives the input values. Layer 2 or the feedback layer represents the computation of the Gaussian values. Layer 3 or the fuzzy layer represents the fuzzy rules. Layer 4 is the output of the RFNN, where each node forms a separate output of the system. The links between layer 3 and layer 4 are formed by the weights [1]. In the following, we will present a modification of the above RFNN principle for the needs of our system for controlling the DC (direct current) motors of a robotic actuator.

III. DESCRIPTION OF A TWO-WHEELED ROBOT CONTROL

In our project, we used a mathematical model of a two-wheeled mobile QuickBot robot [4] with DC motors, which is shown in Fig. 2. The dynamic model was derived analytically and implemented in the MATLAB programming environment, where we performed all the navigation control simulations.

The linear dynamical system of the mobile robot with DC motors has the following form in a matrix notation:

$$\dot{x}(t) = Ax(t) + Bu(t), \quad (1)$$

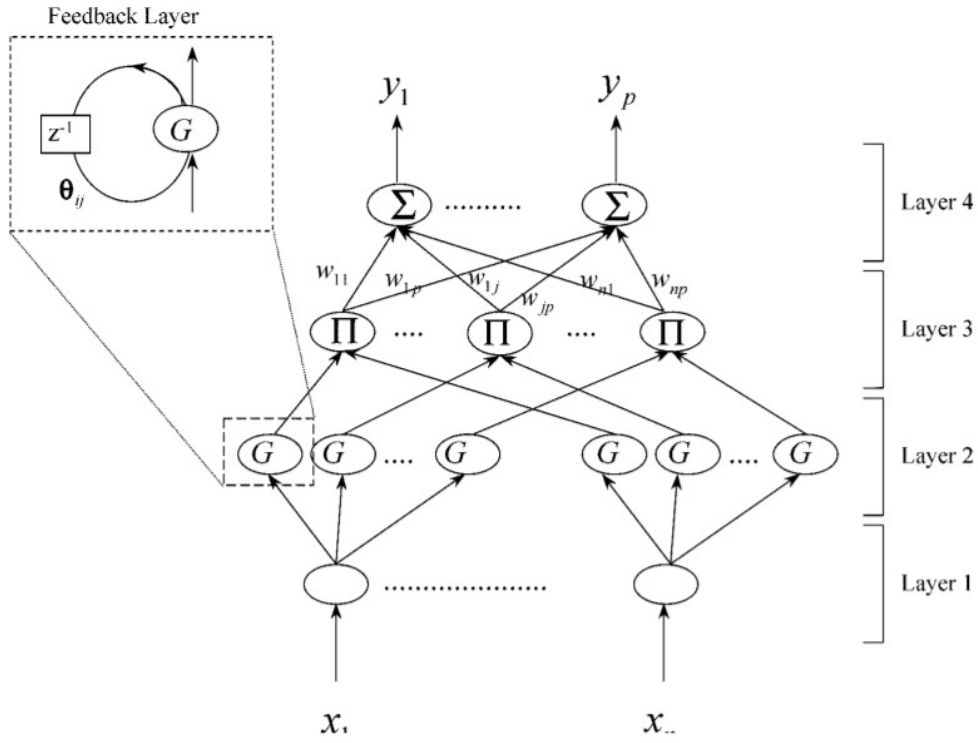


Fig. 1 Topological structure of RFNN [1].

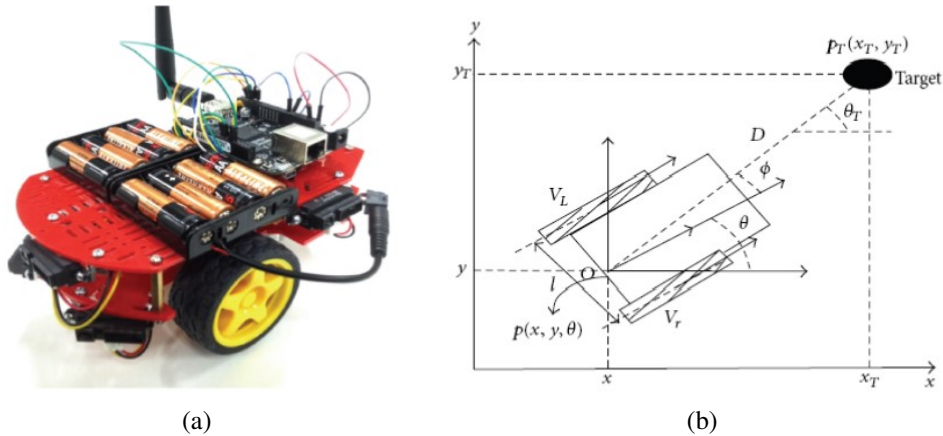


Fig. 2 QuickBot mobile robot and a schematic description of its dynamic model.

where the dynamics matrix A of a mobile robot with DC motors is defined as follows (a detailed explanation of used symbols is in [7]):

$$\begin{bmatrix} -2\frac{Bze}{Mr^2} & 0 & kt\frac{N}{Mr} & kt\frac{N}{Mr} \\ 0 & -2Bze\frac{b^2}{Ir^2} & ktN\frac{b}{Ir} & ktN\frac{b}{Ir} \\ -\frac{60}{2\pi}ke\frac{N}{rLr} & -\frac{60}{2\pi}keN\frac{b}{rLr} & -\frac{Rv}{Lr} & 0 \\ -\frac{60}{2\pi}ke\frac{N}{rLr} & \frac{60}{2\pi}keN\frac{b}{rLr} & 0 & -\frac{Rv}{Lr} \end{bmatrix}. \quad (2)$$

Equation (3) represents a state vector that contains four state variables, the first two being the linear and angular velocity of the robot and the other two being the current flowing through the right and left DC motors (see Fig. 2b).

$$\mathbf{x}^T = [v \quad \omega \quad i_r \quad i_l]. \quad (3)$$

In equation (4), an input vector is defined that contains the voltages at terminals of the right and left motors.

$$\mathbf{u}^T = [u_R \quad u_L]. \quad (4)$$

From the dynamical model, the actual linear and angular velocity can be obtained. In order to know the current position of the robot, a kinematic model is needed, which is described by the relation (5).

$$\dot{\mathbf{q}} = [\dot{x} \ \dot{y} \ \dot{\varphi}] = \begin{bmatrix} \cos \varphi & 0 & \sin \varphi & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} v \\ \omega \end{bmatrix}. \quad (5)$$

The position coordinates are then obtained as the solution of a system of the first order differential equations:

$$\mathbf{q}^T = [x \ y \ \varphi]. \quad (6)$$

To derive the equations of motion, the second order Lagrange equations can be used, where one must first determine the vector n of independent generalized coordinates \mathbf{q} and construct the Lagrange function $L(\mathbf{t})$, which is obtained as the difference of the total kinetic and potential energies of the system. This results in the following n equations of motion:

$$\frac{d}{dt} \left(\frac{\partial L(t)}{\partial \dot{q}_i(t)} \right) - \frac{\partial L(t)}{\partial q_i(t)} + \frac{\partial D(t)}{\partial \dot{q}_i(t)} = Q_i, \quad (7)$$

where Q_i is the input acting on i -th generalized coordinate. A more detailed derivation of the system (7) can be found, e.g., in [8].

Finally, the robot model will be controlled by a controller, whose inputs are a reference position set by the user, the current velocity and the current position, which are obtained directly from the mobile robot. Based on this information, the controller must generate a voltage at the terminals of the mobile robot's motors, that will cause the robot to move towards the desired goal.

IV. SOLUTION DESIGN

We used two controllers connected in series to meet our objective. The first one is a kinematics controller, whose task is to generate a reference velocity based on the position deviations, which are calculated as the difference between the current and reference positions. If there is an obstacle in the path, the system will switch to the avoidance mode and the reference position in the next step is changed to a value such that the robot safely avoids the obstacle. If the obstacle has been safely avoided, the reference position is again overwritten to the original value entered by the user. This controller has been implemented in the form of a Mamdani fuzzy controller.

The second controller is a dynamics controller. Its function is to generate a voltage at the DC motor terminals based on the current and desired speed. This controller was implemented in the form of a neural and fuzzy neural network, which were trained as inverse models for the purpose of further comparison. In other words, this type of serial controller as a whole involves two stages of hybridization. The first stage is the division of the controller into two parts, each of which is implemented by a different computational intelligence method. The second stage is the realization of the controller dynamics using a hybrid fuzzy neural network.

A. Kinematics controller

The role of this controller is to generate a reference velocity (input to the dynamics model) based on positional deviations so that we can minimize these deviations. In the polar coordinate system, the position of the reference point is expressed by the Euclidean distance ρ and the directional rotation deviation α :

$$\rho = \sqrt{\Delta x^2 + \Delta y^2}, \quad (8)$$

$$\alpha = \arctan \frac{\Delta y}{\Delta x}. \quad (9)$$

For its implementation, the Mamdani controller with the center-of-gravity defuzzification was used. The controller has two inputs and two outputs. The inputs are the positional deviations α and ρ and the outputs are the reference angular velocities of the right and left wheels ω_R and ω_L . All fuzzy sets were uniformly distributed to cover the entire universe of input and output variables. The fuzzy linguistic variable ρ is defined on the universe of the Euclidean distance of the robot and the reference position. This linguistic variable is split between three word values *Low*, *Medium* and *High*. The linguistic variable α is defined on the universe of the angular deviation of the robot's pointing from the goal and also takes three word values *Negative*, *Neutral* and *Positive*. The output variables of the kinematics controller ω_R and ω_L are defined on the universe of the angular velocities of the left and right wheels and can take five word values *Very Low*, *Low*, *Medium*, *High* and *Very High*.

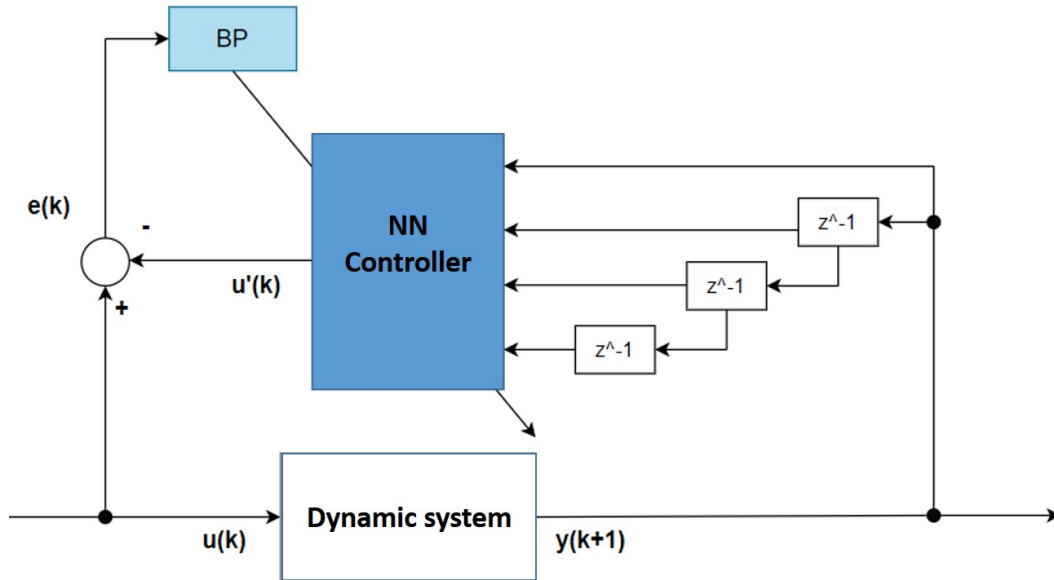


Fig. 3 RFNN in the learning mode.

A rule base that expresses the relationships between these variables in order to approximate the relationships between the inputs and outputs of the controller was designed as follows:

- 1) If $\rho = H$ and $\alpha = +$ Then $\omega_R = VH$ and $\omega_L = M$
- 2) If $\rho = H$ and $\alpha = N$ Then $\omega_R = H$ and $\omega_L = H$
- 3) If $\rho = H$ and $\alpha = -$ Then $\omega_R = M$ and $\omega_L = VH$
- 4) If $\rho = M$ and $\alpha = +$ Then $\omega_R = H$ and $\omega_L = L$
- 5) If $\rho = M$ and $\alpha = N$ Then $\omega_R = M$ and $\omega_L = M$
- 6) If $\rho = M$ and $\alpha = -$ Then $\omega_R = L$ and $\omega_L = H$
- 7) If $\rho = L$ and $\alpha = +$ Then $\omega_R = M$ and $\omega_L = VL$
- 8) If item If $\rho = L$ and $\alpha = N$ Then $\omega_R = L$ and $\omega_L = L$
- 9) If $\rho = L$ and $\alpha = -$ Then $\omega_R = VL$ and $\omega_L = M$

B. Dynamics controller

The dynamics controller has the task of monitoring the reference speed. This means that it must generate such an action (in our case, motor voltages) that the output of the dynamics system follows the reference inputs to the dynamics controller that were generated by the kinematic controller. For this purpose, the RFNN, whose structure was described in the section II, was used and also for comparison purposes the classical MLP with one hidden layer.

Unlike in [1], where this network was trained as a forward model, in this case the network was trained as an inverse model, which was subsequently used as a forward controller. The MLP network was also trained in the same way. Like all neural networks, this one operates in training mode and testing mode, respectively.

In the training mode, the network tries to approximate the inverse dynamic model of the mobile robot, i.e. the outputs from the system represent the inputs to the network and the outputs from the network in turn represent the inputs to the system. In this case, however, the outputs are fed back several steps to the network input in addition to the current system output.

Fig. 3 represents the wiring diagram of the RFNN/NN network in the learning phase. The whole learning process is described in the following steps:

- 1) An input $u(k)$ is provided to the system, which triggers a response $y(k+1)$ in the system.
- 2) The output of the system $y(k+1)$ is fed to the network input, along with the outputs 1–3 steps to the rear.
- 3) Based on the inputs, the network generates an output $u'(k)$ that is an approximation of the real input to the system.
- 4) Based on the difference between the real and the approximated value of the input, the network parameters are adapted using an error backpropagation algorithm.

In the testing mode, the trained inverse model performs the controller function. The block diagram of its circuit is shown in Fig. 4. As in the case of training, the delayed values of the

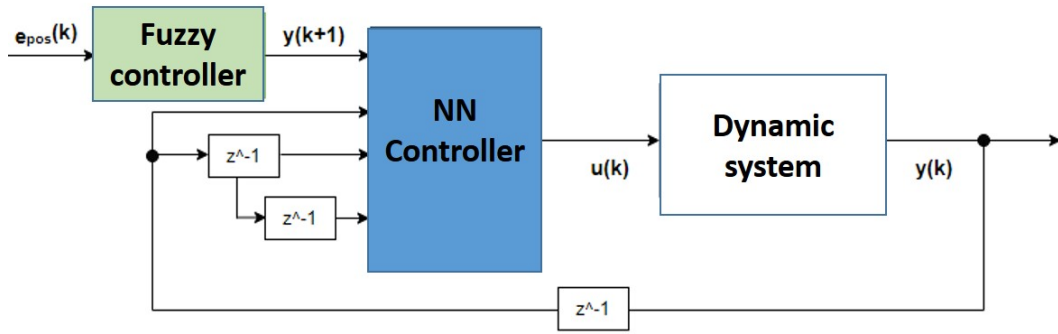


Fig. 4 RFNN in testing phase.

Table 1
 Comparison of the mean squared deviation of MLP and RFNN for different combinations of their parameters.

Number of neurons	Learning parameter	MLP	RFNN
10	0.001	3.257×10^{-5}	3.02×10^{-5}
10	0.002	2.50×10^{-5}	2.19×10^{-5}
10	0.005	5.02×10^{-5}	4.02×10^{-5}
10	0.01	9.20×10^{-5}	8.20×10^{-5}
10	0.05	5.20×10^{-4}	4.33×10^{-4}
11	0.001	5.60×10^{-5}	5.50×10^{-5}
11	0.002	6.66×10^{-4}	5.56×10^{-4}
11	0.005	6.55×10^{-5}	6.25×10^{-5}
11	0.01	7.55×10^{-5}	7.00×10^{-5}
11	0.02	6.34×10^{-5}	6.22×10^{-5}
12	0.001	5.56×10^{-4}	5.45×10^{-4}
12	0.002	7.23×10^{-5}	6.43×10^{-5}
12	0.005	3.31×10^{-4}	2.21×10^{-4}
12	0.01	5.35×10^{-5}	4.33×10^{-5}
12	0.02	3.57×10^{-5}	4.22×10^{-5}
13	0.001	1.36×10^{-5}	1.22×10^{-5}
13	0.002	4.62×10^{-5}	2.22×10^{-5}
13	0.005	1.35×10^{-5}	1.10×10^{-5}
13	0.01	6.67×10^{-5}	5.87×10^{-5}
13	0.02	1.38×10^{-5}	1.23×10^{-5}

outputs from the dynamical system are fed to the controller input. The only difference is the value $y(k + 1)$, which in this case cannot be fed from the dynamic system, since this value is a response to the input $u(k)$. Therefore, this value is obtained from the kinematic fuzzy controller and represents the reference value to which the output value should reach in the next step after applying the input from the inverse model.

V. EXPERIMENTS AND THEIR EVALUATION

The experimental section focuses on determining the accuracy, with which the trained models were able to approximate the inverse dynamics of the mobile robot. Table 1 shows a comparison of the mean square deviation on the test set achieved by MLP and RFNN.

For MLP, the best results were obtained with 13 neurons and a learning parameter of 0.005. Similarly, with the same parameters, the RFNN also achieved the best results. When compared with each other, RFNN comes out slightly better than MLP.

VI. CONCLUSIONS

As experiments confirm, the trained inverse model reliably performs the controller function. All simulations were performed smoothly and the hybrid recurrent fuzzy neural network model made it possible to navigate the two-wheeled robot in a 2D space. From the results, it is clear that better values were observed for the RFNN, which makes this method suitable for the solution.

ACKNOWLEDGMENT

This publication is the result of the APVV grant ENISaC – Edge-Enabled Intelligent Sensing and Computing (APVV-20-0247).

REFERENCES

- [1] C.-H. Lee and C.-C. Teng, "Identification and control of dynamic systems using recurrent fuzzy neural networks," *IEEE Transactions on Fuzzy Systems*, vol. 8, no. 4, pp. 349–366, 2000.
- [2] J. Vaščák, P. Papcun, I. Zolotová, and M. Puheim, "Intelligent space design for robotic and IoT applications," in *2020 Cybernetics Informatics (KI)*, 2020, pp. 1–6.
- [3] K. ichi Funahashi and Y. Nakamura, "Approximation of dynamical systems by continuous time recurrent neural networks," *Neural Networks*, vol. 6, no. 6, pp. 801–806, 1993.
- [4] H. Omrane, M. S. Masmoudi, and M. Masmoudi, "Fuzzy logic based control for autonomous mobile robot navigation," *Comput. Intell. Neurosci.*, vol. 2016, pp. 9 548 482:1–9 548 482:10, 2016.
- [5] R. J. Williams and D. Zipser, "A learning algorithm for continually running fully recurrent neural networks," *Neural Computation*, vol. 1, no. 2, pp. 270–280, 1989.
- [6] B. Šter, "Recurrent neural networks in a mobile robot navigation task," in *Artificial Neural Nets and Genetic Algorithms*, V. Kůrková, R. Neruda, M. Kárný, and N. C. Steele, Eds. Vienna: Springer Vienna, 2001, pp. 169–172.
- [7] P. Guerra, P. Alsina, A. Medeiros, and A. Jr, "Linear modelling and identification of a mobile robot with differential drive," in *Proceedings of the First International Conference on Informatics in Control, Automation and Robotics (ICINCO)*, 01 2004, pp. 263–269.
- [8] R. Siegwart, I. R. Nourbakhsh, and D. Scaramuzza, *Introduction to Autonomous Mobile Robots, Second Edition*, ser. Intelligent robotics and autonomous agents. MIT Press, 2011.

Rozšírenie kolaboratívneho virtuálneho prostredia o podporu interaktívnej gestikulácie a vizualizácie rúk

¹Branislav SOBOTA, ²Marián HUDÁK, ³Matej KOVÁCS, ⁴Štefan KOREČKO, ⁵Miriama MATTOVÁ

^{1,3,4,5}Katedra počítačov a Informatiky, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Letná 9, 04200 Košice, Slovenská republika

²Siemens Healthcare s.r.o., Tr. SNP No. 37, 040 11 Košice, Slovenská republika

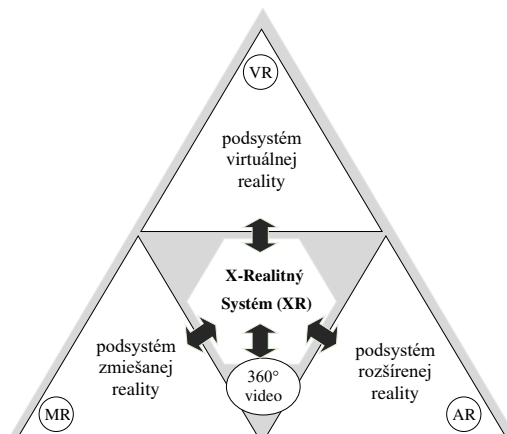
¹branislav.sobota@tuke.sk, ⁴stefan.korecko@tuke.sk

Abstrakt — Tento príspevok sa zaoberá rozšírením systému LIRKIS G-CVE o podporu interaktívnych 3D rúk. LIRKIS G-CVE je vyvíjaný na pracovisku autorov ako implementačná báza pre systémy kolaboratívnej virtuálnej reality. Príspevok obsahuje stručný úvod do problematiky naturálnych rozhraní na báze rozšírenej (XR)/ virtuálnej reality (VR) s fokusáciou na gestikuláciu rúk. Je opísaná základná implementácia rozšírenia a takisto niektoré vykonané experimenty na vzniknutom riešení a ich výsledky. Následne sa na základe výsledkov zhodnotia prínosy riešenia a jeho nedostatky.

Kľúčové slová — virtuálna realita, kolaboratívne virtuálne prostredia, 3D model rúk, naturálne rozhranie, ovládanie gestami rúk

I. ÚVOD

Moderné informačné technológie na čele s technológiami virtuálnej reality (VR) stále viac a viac prenikajú do života každého z nás. Narastá požiadavka neustále interaktívnejších a inteligentnejších (smart) riešení v tejto oblasti. Veľmi progresívnym je použitie práve nových technológií vo vyučovaní alebo tréningu. Jednou z najdynamickejšie rastúcich technologických oblastí, ktoré by mohli úroveň interaktivity citeľne zvýšiť je rozšírená realita (XR, Obr. 1) a v rámci nej virtuálna realita a jej pridružené technológie [1]. Prienik XR/VR technológií do oblasti výučby je v poslednom období badateľný aj z viacerých aplikácií, ako napríklad použitie VR simulátorov pre tréning chirurgov [2] či hasičov [3], kde potreba ovládania rukami používateľa má vyšší význam. Potrebu ich nasadenia v oblasti stavebníctva potvrdzuje aj Luimula a spol. [4]. V tomto smere viacero na pochopenie náročnejších konceptov z oblastí ako sú napríklad fyzika, chémia, biológia, geodézia, stavebníctvo či história, by si však žiadalo omnoho interaktívnejšie sprostredkovanie obsahu. To zo sebou prináša aj neustále vylepšovanie rozhraní týchto systémov.



Obr. 1 Konceptia XR aplikácie platformy

Na pracovisku autorov príspevku (Laboratórium LIRKIS, KPI FEI TU Košice) sa vykonávajú práce, ktoré by mohli pomocou technológií XR v kontexte HCI a HCM-T [5] uľahčiť, zjednodušiť, zrýchliť a sprehľadniť komunikáciu študentov a žiakov s výpočtovou technikou ako aj so svojím okolím. Tento koncept je zaujímavý aj z hľadiska možných dlhodobých cieľov v kontexte rozvoja online vzdelávania, zvýrazneného aj dopadmi predošlej pandemickej situácie. Na strane aplikačného nasadenia môže XR uľahčiť aj komunikáciu opačným smerom t.j. od ich okolia.

Používanie virtuálnych prostredí rozširuje možnosti spolupráce skupiny používateľov v reálnom čase. Zdieľané (kolaboratívne) virtuálne prostredia sú otvorené možnostiam rôznych simulácií, tréningov alebo edukácie a iných typov aktivít. Z pohľadu Flaviána [6], kolaboratívne systémy predstavujú špecifickú architektúru a technologické vybavenie pre možnosti vytvárania kolaboratívnych aktivít. Autori [7] prízvukujú potrebu fyzickej prítomnosti používateľov alebo ich virtuálnej podoby vo forme avatarov v kolaboratívnych prostrediach. Najčastejšie kolaboratívne prostredia na tejto úrovni sú systémy založené na a) *virtuálnych jaskyniach* a systémy založené na b) *virtuálno-reálnych/dátových prilbách* ako formy kolaboratívnych prostredí dostupných pre širokú skupinu používateľov.

Jedným z hlavných problémov kolaborácie pri komunikácii používateľov vo virtuálnom prostredí najmä pre systémy založené na VR/dátových prilbách je potreba vizualizovania gestikulácie rúk (gestá rúk sú v týchto systémoch považované častokrát ako kľúčový vstupný prvok). Podpora prirodzenej gestikulácie rúk je v prípade používania VR/dátových prilb technologicky možná, zároveň komplikovaná. Pohltím vizuálneho vnemu používateľa virtuálnym priestorom nie je možné interagovať s objektmi scény podobne ako v systémoch VR jaskýň, kde používateľ vidí prirodzene fyzicky dostupné vlastné ruky. Z tohto dôvodu sú frekvencovane používané technológie s podporou virtuálnych dotykových ovládačov, ktoré sprostredkujú snímanie kostrového modelu ruky. Pohyb kostrového modelu je následne prenášaný do virtuálneho prostredia, v ktorom používateľ vidí vlastné ruky ako virtuálne. Prenášanie kostrového modelu gestikulácie rúk je súčasne realizované jednoduchými modelmi. Dôvodom je potreba dosahovania primeranej záťaže na výkon výpočtového systému, prijímanie signálov zo zariadenia umiestneného na ruke a prenášanie výpočtov do virtuálneho prostredia. Navýšením počtu používateľov narastajú požiadavky na spracovávanie gestikulačných vzorov a rozhraní cieľovým systémom. Každú z transformácií prstov používateľov je potrebné vizuálne spracovať a prenášať do virtuálneho prostredia [8]. Pre zníženie nárokov na výpočtový výkon a prenášaný objem dát od každého používateľa sa spravidla eliminuje spracovávanie detailov skupiny používateľov. Každý používateľ má k dispozícii detailné zobrazenie len vlastných virtuálnych objektov rúk. Z opačného pohľadu ruky ostatných používateľov sú viditeľné len v podobe jednoduchých modelov. Týmto vzniká otázka riešenia pre distribúciu veľkého objemu údajov v reálnom čase, bez vysokých latencií a s vhodnou reprezentáciou prostredníctvom sémanticky jednoduchých trojrozmerných modelov. VR/dátové prilby z hľadiska ceny aj dostupnosti sú v zmysle širšieho nasadenia vhodným riešením nakoľko je možné aj použitie VR prilb na báze smartfónu. Takto koncipované kolaboratívne prostredia by mohli byť výborným základom pre riešenie projektu. Systém LIRKIS G-CVE [9], vyvíjaný na pracovisku autorov, navyše svojou implementačnou platformou (webové technológie) ďalej rozširuje aplikačnú bázu a znižuje cenu celkového výstupu. Rozšírením LIRKIS G-CVE o podporu interaktívnych 3D rúk je obsahom tohto článku.

Samotný systém LIRKIS G-CVE používa rámec Aframe a je zameraný na interakciu viacerých používateľov v jednej virtuálnej scéne využitím webového prehliadača. Takéto virtuálne prostredia, kde môžu spolu interagovať viacerí používatelia, sa nazývajú kolaboratívne virtuálne prostredia (CVE). Ako je uvedené v [10], webové kolaboratívne virtuálne prostredia majú mnoho silných stránok pre zvýšenie použiteľnosti a dopytu pre virtuálnu realitu. Medzi základné benefity využitia webových CVE sú tieto:

- rýchly vývoj interaktívnych a komunikačných rozhraní,
- podpora viacerých platforiem a hardvéru virtuálnej reality,
- plný prístup používateľov bez ohľadu na geografickú vzdialenosť medzi nimi,
- nákladovo efektívnejšie riešenie ako štandardné aplikácie virtuálnej reality.

II. INTERAKCIA S OBJEKTAMI V SYSTÉME LIRKIS G-CVE

V systéme LIRKIS G-CVE boli zatiaľ riešené interakcie pomocou existujúcich komponentov. Jedno z riešení je možné uviesť ako príklad. Toto riešenie obsahuje malú miestnosť spolu so stolíkom a s objektami, ktoré podporujú interakciu. Teda tieto objekty je možné uchopiť a manipulovať s nimi. Riešenie podporuje aj zdieľanie a avatarov jednotlivých používateľov,

takže každý používateľ dokáže sledovať polohu druhého používateľa vo virtuálnom priestore v reálnom čase. Jedným z hlavných nedostatkov tohto riešenia je vizuálna stránka interakcie s objektami. Ako reprezentácia polohy ruky používateľa je iba model ovládača. Samotné ovládače, ktoré iba lietajú vo vzduchu narúšajú používateľovi pohltie do zážitku VR. Tento model nijak nie je pripojený k avatarovi a tento model ani nie je zdieľaný medzi ostatných používateľov vo virtuálnej scéne. To znamená že ostatní používatelia nevidia s čím interaguje používateľ a kde sa jeho ruky nachádzajú. Preto cieľom pri vytváraní opisovaného riešenia bolo doplnenie týchto nedostatkov najmä vytvorením 3D modelu rúk, ktorý je prirodzený a je pre používateľa realistickejší, čím sa zvýši jeho pohltie do VR. Navyše bola snaha zabezpečiť, že tento model virtuálnych rúk bude zdieľaný a aj ostatní používatelia budú vidieť, kde sa ruky daného používateľa nachádzajú v reálnom čase. Originálna virtuálna scéna bez popisovaného rozšírenia je na obrázku Obr. 2 (poloha rúk je vizualizovaná modelmi ovládačov).



Obr. 2 Virtuálna scéna pred využitím rozšírenia virtuálneho modelu rúk

III. PRINCÍPY VYTVÁRANIA MODELOV A ICH POHYBU

Vo virtuálnom prostredí by mali objekty byť čo najviac prirodzené, aby pôsobili vierohodne a v čo najväčšej miere oklamali naše zmysly. Zvieratá, rastliny a samozrejme človek sú neodmysliteľnou súčasťou takéhoto prostredia. Systémy XR/VR sa počas svojho vývoja zaoberali aj vizualizáciou stále zložitejších scén, ktoré sa čoraz viac približovali realite. Preto sa stalo nevyhnutným aj spomínané zobrazovanie objektov. Nakoľko tieto objekty nie sú triviálne, ich modelovanie a simulácia nie je jednoduchá. V minulosti sa pre tieto účely využívala najmä aproximácia objektov pomocou jednoduchých geometrických telies ako kváder, kužeľ a guľa [11].

Z hľadiska modelovania osôb sa v súčasnosti využívajú metódy modelovania pomocou primitív, kriviek, polygonálne a sculp modelovanie. V prípade sculp modelovania sa definujú tri typy modelov organizmov. Postupnou aplikáciou jednotlivých modelov získame zobrazenie organizmu, resp. osoby. Prvým modelom je skeletálny model. Skeletálny model tvorí sústava kostí a kĺbov, na ktorý sa aplikuje muskulárny model predstavujúci svalstvo organizmu. Nakoniec sa aplikuje povrchový model, ktorý predstavuje povrch organizmu – najčastejšie kožu.

Dôveryhodné zobrazenie modelu osoby nezahŕňa len vizuálny virtuálny dojem. Dôležitou súčasťou modelu je jeho dynamika a interakcia s prostredím. Dynamika predstavuje pohyb objektu, resp. jeho častí a treba ho definovať ako zmenu stavu v čase. Zmena stavu objektu musí byť vyjadrená voči tzv. referenčnému bodu. Zložitosť opisovaného pohybu narastá zvyšovaním počtu častí objektu, ktoré sú súčasťou pohybu a vytvárajú medzi sebou vzťahy [11].

Simuláciu pohybu objektu umožňuje zmena stavu jednotlivých častí, ktoré tvoria spolu štruktúru. V prípade modelu osoby je takouto štruktúrou skeletálny model, a teda kostra človeka. Ako bolo spomínané, skeletálny model pozostáva z kostí a kĺbov. Práve kĺby sú časťou štruktúry, vďaka ktorej jednotlivé časti kostry, ktorými sú kosti, môžu zmeniť svoj stav. Z toho dôvodu je pre pohyb potrebné zadefinovať kĺbovú štruktúru modelu. Zmena stavu je vyjadrená voči referenčnému bodu. V kĺbovej štruktúre takýto bod predstavuje tzv. statický kĺb nazývaný aj báza. Kĺb, ktorého zmenu voči báze sledujeme, je dynamický kĺb, nazývaný aj koncový efektor. Medzi bázou a koncovým efektorom môže existovať viacero kĺbov, prípadne žiadny. Kĺby od statického kĺbu, bázy, po koncový efektor, tvoria *kinematický reťazec*.

Rekonštrukcia pohybu rúk do virtuálneho sveta nie je jednoduchá. Na určenie pohybu je potrebné si definovať kĺbovú štruktúru. Bázou, v prípade vytvárania virtuálnych rúk, bude ramenný kĺb, ktorý je v tomto prípade statický a koncovým efektorom bude dlaň ruky. Z pohľadu, či potrebujeme vedieť koncovú pozíciu koncového efektora alebo veľkosti uhlov všetkých kĺbov kĺbovej štruktúry, vieme kinematiku takéhoto modelu rozdeliť na priamu a inverznú. Pri priamej kinematike platí funkcia (1), ktorá popisuje nájdenie polohy a orientácie koncového efektora pre kinematický reťazec.

$$f(\text{uhly_kĺbov}) = \text{poloha_koncového_efektora} \quad (1)$$

Pri inverznej kinematike, platí funkcia, ktorá je opačná k funkcii priamej kinematiky (2). V tomto prípade sa určujú veľkosti uhlov jednotlivých kĺbov kinematického reťazca vzhľadom na polohu koncového efektora.

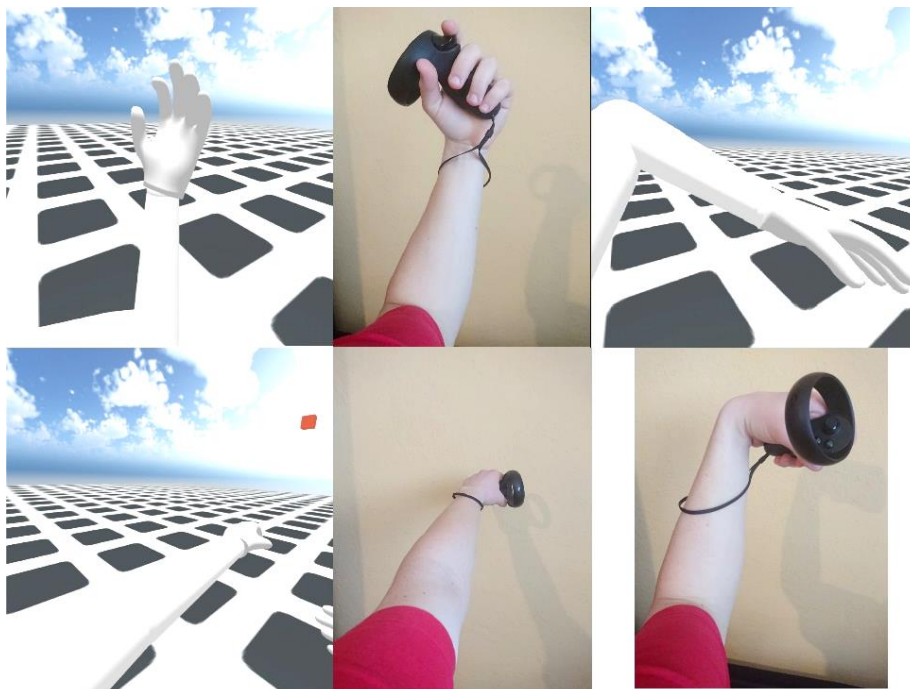
$$f^{-1}(\text{poloha_koncového_efektora}) = \text{uhly_kĺbov} \quad (2)$$

Pre riešenie modelovania virtuálnych rúk bude teda výhodnejšie použiť inverznú kinematiku z dôvodu, že pri použití virtuálnych prílb pozícia ovládačov určuje pozíciu koncového efektora, podľa ktorého vieme zistiť veľkosti uhlov pre daný stav modelu. Problémom však pre inverznú kinematiku zostáva určiť veľkosť uhlov kĺbov tak, aby koncový efektor dosiahol očakávané pozície. Pre riešenie tohto problému je možné využiť dva rôzne typy metód [11]:

- Analytické metódy, inak nazývané aj exaktné metódy, ktorých cieľom je nájsť vyjadrenie matematickej funkcie f^{-1} . Práve preto sú analytické metódy také presné aj keď niekedy nie rýchle alebo implementačne zložitejšie. Ako príklad je možné uviesť inverznú kinematiku s využitím konformnej geometrickej algebry.
- Numerické alebo aj iteratívne metódy, neposkytujú takú presnosť ako analytické metódy ale v mnohých prípadoch umožnia jednoduchšiu implementáciu a tým vyššiu rýchlosť. V prostredí VR sa neočakáva úplná presnosť, nakoľko ľudské oko nedokáže tieto nepresnosti zachytiť. V niektorých prípadoch využitia sú tieto metódy preferované namiesto analytických metód. Príkladom je metóda cyklického približovania súradníc: CCD – Cyclic Coordinate Descent.

IV. OPIS VYTVORENÉHO RIEŠENIA

Počas implementácie vznikli dve riešenia. Obe používali existujúce komponenty použitého rámca A-frame: *super-hands*, *hand-controls* a *look-at*. Tieto existujúce komponenty zabezpečujú funkcie ako interakciu s objektami virtuálnej scény, model prstov a dlane a rotáciu objektov, z ktorých sa celkový virtuálny model rúk skladá. Prvé riešenie využívalo iba vlastnosť dedenia rámca A-frame, čím sa zabezpečil plynulý pohyb modelu. Problémom tohto riešenia bola strata voľnosti v kĺbe zápästia. Niektoré polohy virtuálneho modelu ruky využitím tohto riešenia sú na obrázku Obr. 3.



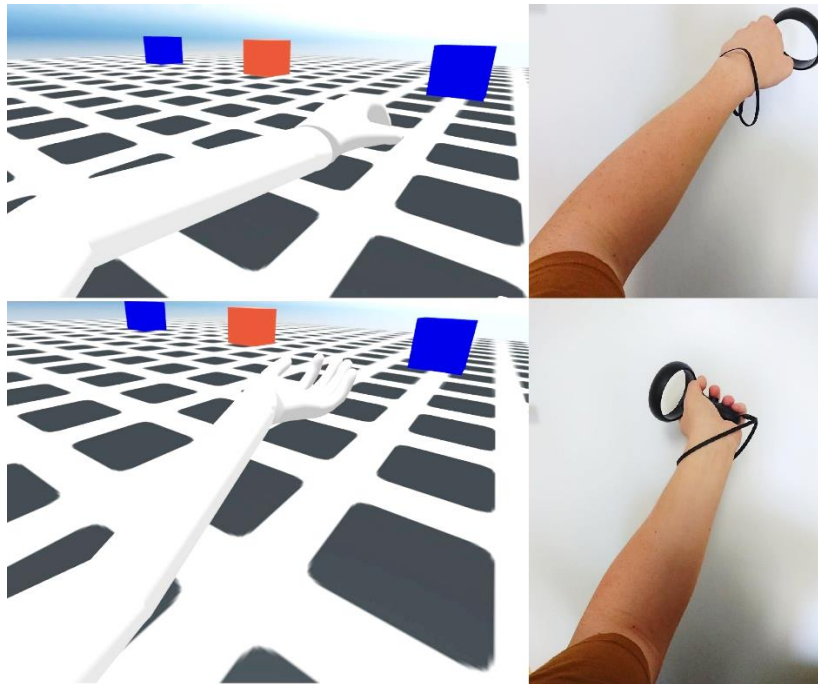
Obr. 3 Polohy modelu virtuálnej ruky a ruky používateľa v reálnom svete

Ďalším riešením bola implementácia vlastných komponentov. Tieto komponenty reprezentovali jednotlivé kĺby modelu. Najdôležitejšími komponentami boli komponent zápästia a komponent lakt'ového kĺbu. Funkciou samotného komponentu zápästia bolo odčítanie rotácie ruky používateľa, teda rotácie ovládača. Z tejto odčítanej rotácie na základe viacerých podmienok, ktoré zabezpečujú určité obmedzenie pohybu zápästia sa nastaví objektu zápästia

správna rotácia. Po správnom nastavení rotácie objektu zápästia dokáže komponent lakt'ového kĺbu vypočítať pozíciu lakt'ového kĺbu vo virtuálnej scéne. Komponent lakt'ového kĺbu postupným výpočtom získa pozíciu lakt'ového kĺbu:

1. Ako prvým krokom výpočtu je nájdenie stredového bodu medzi ramenným kĺbom a zápästím.
2. Následne je potrebné určiť vzdialenosť lakt'ového kĺbu od stredového bodu.
3. Potom je z tejto vzdialenosti možné určiť vektor od stredového bodu.
4. Následne na tento vektor sa aplikujú dva quaterniony, ktoré upravujú rotáciu tohto vektora okolo stredového bodu. Touto úpravou rotácie je možné nakoniec dostať správnu pozíciu lakt'ového kĺbu vo virtuálnej scéne.

Niektoré polohy virtuálneho modelu rúk využívajúceho vlastné komponenty sú na nasledujúcom obrázku Obr. 4.



Obr. 4 Polohy virtuálnej ruky a ruky používateľa v reálnom svete

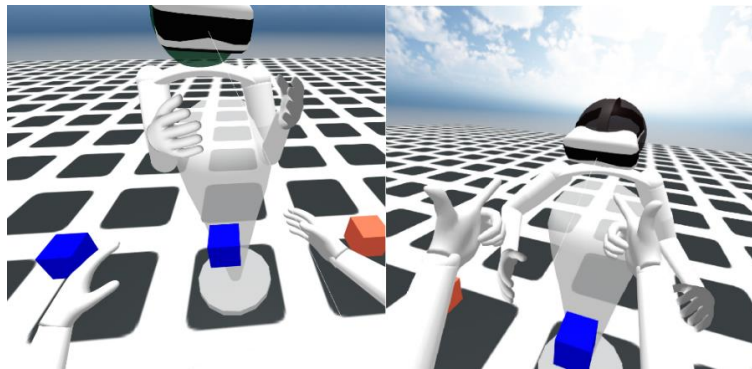
V. EXPERIMENTÁLNE OVERENIE RIEŠENIA

A. Overenie správania sa modelu rúk

Pre overenie správania modelu virtuálnych rúk sa použila základná scéna, v ktorej bol model vytvorený. V tejto scéne bol jeden objekt, ktorý bol nastavený ako interaktívny, čiže používateľ s ním mal vedieť manipulovať. Používateľ počas experimentu bol schopný objekt chytiť a presúvať vo virtuálnej scéne. A teda aj interakcia s objektami bola možná a fungovala podľa očakávaní. Počas experimentov sa však odhalili aj nepresnosti v riešení kĺbov modelu virtuálnych rúk. V určitých prípadoch keď rotácia používateľa vo svete bola viac ako 180° nastával problém v podmienkach kĺbu zápästia. V komponente pre správanie tohto kĺbu sú podmienky využívajúce uhly ruky používateľa. V prípade keď sa používateľ otočí o 180° sa tieto podmienky nevyhodnotia správne. Implementácia na základe týchto experimentov bola upravená, ale tento problém sa nepodarilo odstrániť úplne.

B. Chyby zdieľania zistené pri experimentoch

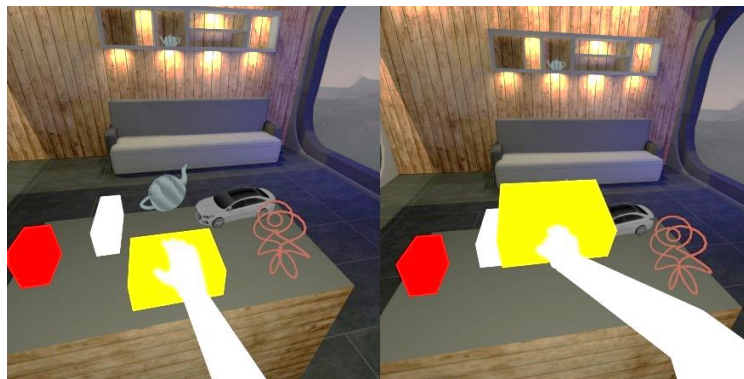
Experiment odhalil aj problém so zdieľaním ruky používateľa. Problém spočíval v tom, že entity ovládačov využívajúce komponent *hand-controls* svoj model získavajú z komponentu. Toto zabraňuje zdieľaniu modelu ruky. Riešenie pre tento problém bolo len čiastočné. Nasledovne sa šablóna zdieľania upravila tak, aby sa nezdieľala entita s komponentom *hand-controls*, ale iba modely rúk. To však zabraňuje zdieľaniu gest rúk, teda používateľ u seba vidí vykonávané gestá, ale druhý používateľ v scéne vidí iba statický model ruky druhého používateľa. Pohľady z perspektívy oboch používateľov sú ilustrované na obrázku Obr. 5. Ak by parameter lokálneho zdieľania bol nastavený na hodnotu *pravdivý*, používateľ by videl aj model komponentu *hand-controls* a aj model ktorý sa zdieľa pomocou komponentu *networked*. Aj toto riešenie je však už výrazným pokrokom oproti pôvodnému stavu.



Obr. 5 Pohľad na zdieľané ruky (vľavo), pohľad u používateľa lokálne (vpravo)

C. Overenie použiteľnosti riešenia v inom projekte

Druhý experiment bol zameraný a integráciu modelu do už existujúcej aplikácie. V tomto prípade úlohou bolo vyskúšať, ako je možné implementáciu preniesť do existujúcej aplikácie. Vybraná aplikácia obsahovala niekoľko objektov, s ktorými používateľ vedel interagovať vo virtuálnej scéne s virtuálnymi ovládačmi. Tieto ovládače boli v aplikácii nahradené modelom virtuálnych rúk. Pridanie modelu do virtuálnej scény bolo jednoduché a na základe diagramu scény sa podarilo rýchlo určiť kde sa majú potrebné entity pridať do scény aby fungovali správne. Vyskytol sa problém pri interakcii rúk a objektov. Riešenie tohto problému bolo však jednoduché. Namiesto navrhovaného komponentu *super-hands* v implementácii modelu virtuálnych rúk, bol tento komponent nahradený upraveným komponentom *super-hands* z existujúcej aplikácie. Výsledná interakcia rúk (Obr. 6) a objektov ostala tým pádom zachovaná a funkčná.



Obr. 6 Interakcia uchopenia objektu v už existujúcom projekte

D. Meranie náročnosti riešenia na hardvér

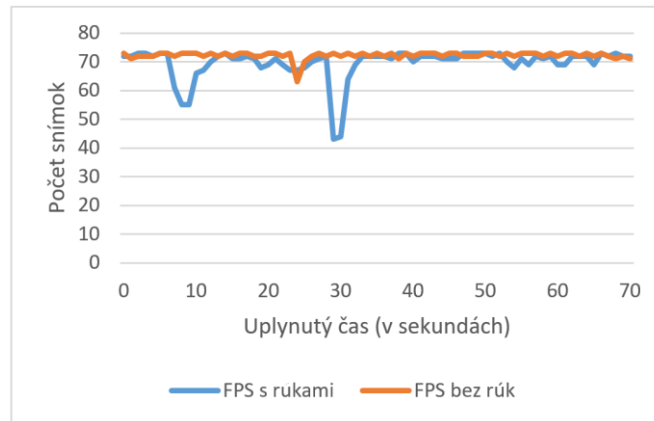
Pre VR je veľmi dôležité aby celá virtuálna scéna bola zobrazovaná používateľovi plynule a bez prerušenia. Preto je potrebné ohodnotiť vytvorenú implementáciu a jej vplyv na celkový výkon. Merania sa vykonali na dvoch rovnakých virtuálnych scénach. Rozdielom v scénach bola prítomnosť modelu virtuálnych rúk. Počas meraní sa zaznamenávali dve hodnoty, počet snímkov za sekundu (FPS) a záťaž procesora v percentách. Priemerné hodnoty týchto metrik sú uvedené v tabuľke (Tab. 1).

Tab. 1 Priemerné namerané hodnoty

Priemerné namerané hodnoty	
FPS s rozšírením podpory rúk	70 FPS
FPS bez rozšírenia podpory rúk	72 FPS
Zaťaž CPU s rozšírením podpory rúk	72 %
Zaťaž CPU bez rozšírenia podpory rúk	38 %

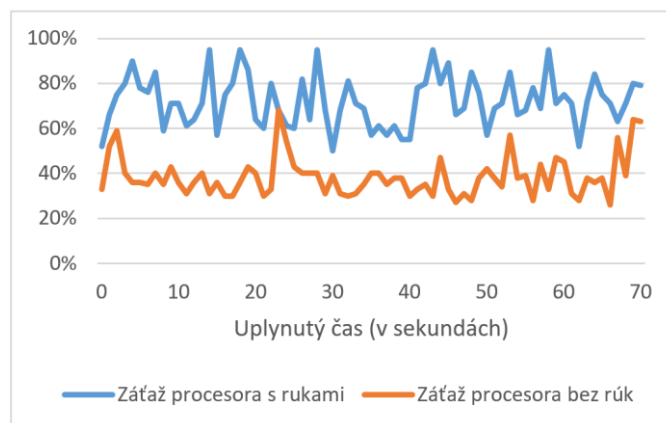
Pri meraní týchto hodnôt sa bral do úvahy rovnaký časový úsek sedemdesiat sekúnd. Počas priebehu merania sa vo virtuálnych scénach vykonali dva rozličné scenáre. V prvom prípade to bol scenár, kedy sa využíval model virtuálnych rúk. Teda používateľ sa vo svete pohyboval, hýbal rukami a manipuloval s objektami. V druhom prípade to bol scenár keď scéna neobsahovala model virtuálnych rúk a používateľ sa iba pohyboval vo virtuálnom svete bez interakcií s objektami.

V prvom porovnaní týchto dvoch scenárov sa porovnal počet snímkov. Porovnaním týchto dvoch hodnôt je možné určiť aký vplyv má implementácia na plynulosť zobrazenia virtuálnej scény pre používateľa. Z grafu (Obr. 7) je vidieť, že tento vplyv nebol kritický. Počas sedemdesiatich sekúnd bol počet snímkov pri využití modelu virtuálnych rúk nižší ako bez jeho využitia. Spolu iba desať sekúnd. Aj tieto poklesy počtu snímkov boli len v dvoch prípadoch väčšie ako 10 snímkov. Pri meraní tieto poklesy používateľ nezaznamenal, jeho zážitok nebol ovplyvnený.



Obr. 7 Graf porovnávajúci počet snímkov za sekundu

Meranie zahŕňalo aj meranie záťaže CPU. Z grafu (Obr. 8) je vidieť, že záťaž sa zvýšila o tridsať percent. Hlavnou príčinou je zahrnutie modelu v scéne, ako aj potrebné výpočty pre určenie polohy modelu. Pre zníženie tohto rozdielu je potrebná ďalšia optimalizácia kódu.



Obr. 8 Záťaž procesora počas merania

VI. ZÁVER

Vytvorené rozšírenie LIRKIS G-CVE o interaktívny model 3D rúk ponúka zvýšenie pohltienia používateľa do zážitku virtuálnej reality. Pridáva vo virtuálnej scéne do veľkej miery realistický model virtuálnych rúk, ktorý kopíruje polohu ruky používateľa v reálnom svete. Taktiež dodáva zdieľanie celého modelu virtuálnych rúk v reálnom čase. Aj keď je riešenie správne a funkčné, experimenty odhalili niektoré nedostatky modelu. Tieto nedostatky je možné ďalšou úpravou riešenia neskôr odstrániť. Medzi tieto úpravy, teda vylepšenia modelu je potrebné zahrnúť: úpravu podmienok v komponente zápästia modelu virtuálnych rúk, lepšiu optimalizáciu výpočtov v zdrojovom kóde pre zníženie záťaže systému, tak aj možnú úpravu modelov jednotlivých častí ruky a zníženie počtu trojuholníkov, z ktorých sa tieto modely skladajú.

Ako už bolo spomenuté, na pracovisku autorov sa vykonávajú práce, ktoré by mohli pomocou technológií XR/VR v kontexte HCI a HCM-T uľahčiť, zjednodušiť, zrýchliť a sprehľadniť komunikáciu používateľov s výpočtovou technikou ako aj so svojím okolím. Tento koncept je zaujímavý aj z hľadiska možných dlhodobých cieľov v kontexte rozvoja online vzdelávania. Rozšírenie o podporu interaktívnej gestikulácie a vizualizácie rúk tak bude snád' ďalším krokom k rozvoju aplikácie moderných technológií aj do tejto oblasti.

POĎAKOVANIE

Táto práca vznikla s podporou projektu KEGA č. 048TUKE-4/2022: „Technológie zdieľanej virtuálnej reality vo vzdelávacom procese“.

REFERENCIE

- [1] Sobota, B.; Korečko, Š.; Hrozek, F.: On building an object-oriented parallel virtual reality system; In: Central European Journal of Computer Science. Vol. 2, no. 3 (2012), p. 261-271. - ISSN 2081-9935
- [2] Mishra, K.; Boland, M.V.; Woreta, F.A. Incorporating a virtual curriculum into ophthalmology education in the coronavirus disease-2019 era. In: Curr. Opin. Ophthalmol. 2020, 31, 380–385.
- [3] Wijkmark, C.H.; Fankvist, S.; Heldal, I.; Metallinou, M.M. Remote Virtual Simulation for Incident Commanders: Opportunities and Possibilities. In: Proceedings of the 2020 11th IEEE International Conference on Cognitive Infocommunications (2020); pp. 445–452.
- [4] Luimula, M.; Linder, M.; Pieskä, S.; Laimio, E.; Lähde, T.; Porraro, P. Unlimited Safety Productivity—A Finnish Perspective Using Virtual Learning Methods to Improve Quality and Productivity in the Construction Industry. In: Proceedings of the 2020 11th IEEE International Conference on Cognitive Infocommunications (2020); pp. 263–266.
- [5] Sinha, G.; Shahi, R.; Shankar, M.: Human Computer Interaction. In: Emerging Trends in Engineering and Technology (ICETET), 2010 3rd International Conference on. IEEE, 2010. p. 1-4
- [6] Flavián, C., Ibáñez-Sánchez, S., & Orús, C. (2018). The impact of virtual, augmented and mixed reality technologies on the customer experience. Journal of Business Research.
- [7] Gonzalez-Franco, M., Hall, M., Hansen, D., Jones, K., Hannah, P., & Bermell-Garcia, P. (2015, March). Framework for remote collaborative interaction in virtual environments based on proximity. In: 3D User Interfaces (3DUI), 2015 IEEE Symposium on (pp. 153-154). IEEE.
- [8] Le Chénéchal, M., Duval, T., Gouranton, V., Royan, J., & Arnaldi, B. (2016, March). Vishnu: Virtual immersive support for helping users—an interaction paradigm for collaborative remote guiding in mixed reality. In: 3DCVE 2016: International Workshop on Collaborative Virtual Environments (pp. 1-5). IEEE.
- [9] Hudák, M.; Korečko, Š.; Sobota, B. : Enhancing Team Interaction and Crossplatform Access in Web-based Collaborative Virtual Environments, In: Proceedings of 2019 IEEE 15th International Scientific Conference on Informatics, IEEE, 2019, pp. 160-16.
- [10] Hudák, M.; Sobota, B.; Korečko Š. and Sivý M.: "Fully Immersive Web-Based Collaborative Virtual Environment for Upper Limb Rehabilitation Purposes," 2020 18th International Conference on Emerging eLearning Technologies and Applications (ICETA), 2020, pp. 194-199, doi: 10.1109/ICETA51985.2020.9379239.
- [11] Sobota, B.; Hrozek, F.: Systémy Virtuálnej Reality, vydanie: prvé, vydavateľ: Technická univerzita v Košiciach, 2015, ISBN: 978-80-553-1970-4

Skúmanie dostupnosti a chybovosti FSO systému vplyvom hmly

¹Maroš Lapčák, ²Luboš Ovseník, ³Jakub Oravec, ⁴Norbert Zdravecký, ⁵Samuel ANDREJČÍK

^{1,2,3,4,5} Katedra elektroniky a multimediálnych telekomunikácií, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

¹maros.lapcak@tuke.sk, ²lubos.ovsenik@tuke.sk, ³jakub.oravec@tuke.sk,
⁴norbert.zdravecky@tuke.sk, ⁵samuel.andrejcek@tuke.sk

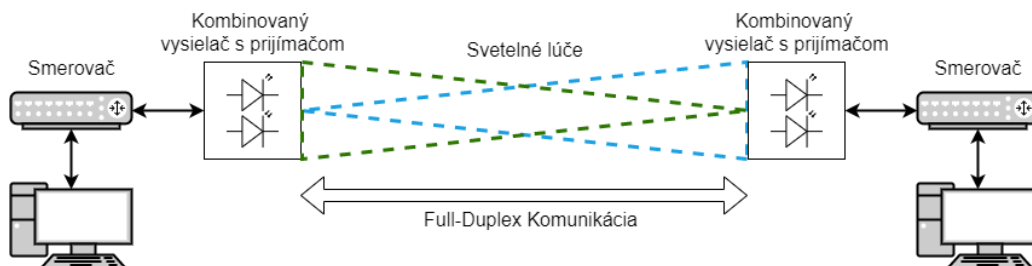
Abstrakt — Táto publikácia sa zaoberá skúmaním dostupnosti a chybovosti FSO systému na základe útlmu spôsobeného hmlou. Z teórie vieme, že na FSO systémy má najväčší vplyv spomedzi atmosférických javov práve hmla. V dôsledku toho sme sa rozhodli odsimulovať a analyzovať tento vplyv v programe OptiSystem a následne porovnať výsledky a zistiť, pri ktorej hodnote útlmu už nie je možné prenášať svetelné lúče z jednej FSO hlavice na druhú.

Kľúčové slová — dostupnosť FSO systému, FSO systém, hmla, chybovosť, útlm spôsobený hmlou

I. ÚVOD

Free-space optics (FSO), nazývaná aj ako optická komunikácia vo voľnom priestore, označuje prenos modulovaných viditeľných alebo infračervených (IR) lúčov cez atmosféru za účelom získania širokopásmovej komunikácie. Najčastejšie používané sú laserové lúče, no k ich funkčnosti postačujú aj slabšie zdroje svetla. Sú to napríklad svetlo emitujúce diódy (LED) alebo diódy emitujúce infračervené spektrum svetla (IRED) [1].

Teória FSO je takmer rovnaká ako teória pre prenos z optických vlákien. Rozdielom medzi nimi je, že energetický lúč je kolimovaný a vysielaný čistým vzduchom alebo priestorom zo zdroja do cieľa a nie je vedený optickým vláknom. Ak zdroj energie neprodukuje dostatočne rovnobežný lúč na prekonanie požadovanej vzdialenosti, je možné vykonať kolimáciu pomocou šošoviek. Zdrojom svetelného žiarenia sú viditeľné alebo infračervené lúče modulované s dátami, ktoré sa majú prenášať. V cieľovom mieste je lúč zachytený fotodetektorom, dáta sú extrahované z viditeľného alebo infračerveného lúča (demodulované) a výsledný signál je zosilnený a odoslaný do hardvéru [1]. Princíp fungovania FSO systému je znázornený na Obr. 1.



Obr. 1 Príklad fungovania FSO systému.

Systémy FSO môžu fungovať na vzdialenosti niekoľkých kilometrov. Pokiaľ je medzi zdrojom a cieľom zreteľný priamy výhľad, je komunikácia teoreticky možná. Aj keď nie je žiadna priama viditeľnosť, na odrážanie energie možno použiť strategicky umiestnené zrkadlá. Svetelné lúče môžu taktiež prechádzať cez sklenené okná s malým alebo žiadnym útlmom v prípade, že okná sú udržiavané čisté [1].

II. VPLYV HMLY NA FSO SYSTÉM

Častice oparu a hmly pozostávajú zo zmesi veľmi jemných kvapiek vody alebo ľadu blízko povrchu Zeme. Svetlo je týmito časticami rozptýlené a tým sa znižuje viditeľnosť. Hmla je charakterizovaná viditeľnosťou kratšou ako 1 km a relatívnou vlhkosťou, ktorá dosahuje úroveň

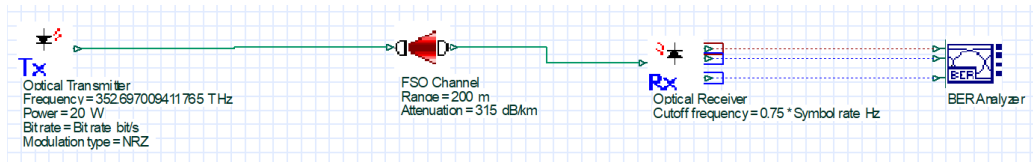
nasýtenia (100 %). Niektoré parametre, ako je distribúcia veľkosti častíc, obsah kvapalnej vody, teplota a vlhkosť atď., definujú hmlu [2]. Najdôležitejším parametrom je distribúcia veľkosti častíc použitá na modelovanie modifikovanej gama distribúcie. Vlastnosti hmly sa môžu líšiť od jednej hmly k druhej alebo dokonca počas tej istej hmly. To závisí od určitých premenných, ako je ročné obdobie, lokalita, atď.. Preto na vysvetlenie každej udalosti samostatne je dôležité poznať parametre hmly, najmä distribúciu veľkosti častíc [3]. Vo všeobecnosti je veľkosť častice hmly porovnateľná s vlnovou dĺžkou signálu FSO, takže pre FSO spojenia spôsobuje veľký útlm. Útlm v hustej hmle presahuje 480 dB/km a v miernej hmle 130 dB/km. Dosah viditeľnosti sa očividne znižuje so zvyšujúcou sa koncentráciou hmly vo vzduchu. Na predpovedanie útlmu je potrebné určiť veľkosť a obsah vody v časticiach hmly. Tieto znalosti je však ťažké dosiahnuť a nie sú vždy dostupné na mieste inštalácie spojenia FSO. Výskumníci preto navrhli analytické modely, ktoré sa spoliehajú na údaje o viditeľnosti, ktoré sú bežne dostupné z meteorologických staníc v mestách. [4, 5].

Tabuľka 1 Medzinárodné hodnoty útlmov hmly

Medzinárodné kódy viditeľnosti			
Atmosférické podmienky	Poveternostné zložky [mm/h]	Viditeľnosť [m]	Útlm [dB/km]
Hustá hmla		50	315
Silná hmla		200	75
Mierna hmla		500	28,9
Slabá hmla	Búrka 100	770	18,3
Veľmi slabá hmla		1000	13,8
	Sneh		
	Silný dážď 25	1900	6,9
Jemná hmla		2000	6,6
	Priemerný dážď 12,5	2800	4,6
Veľmi jemná hmla		4000	3,1
	Slabý dážď 2,5	5900	2
		10000	1,1
Čistý vzduch	Mrholenie 0,25	18100	0,6
		20000	0,54
Veľmi čistý vzduch		23000	0,47
		50000	0,19

III. DOSTUPNOSŤ A CHYBOVOSŤ FSO LINKY VPLYVOM HMLY

Keďže útlm hmly patrí medzi najväčšie vplyvy atmosférického kanála na FSO linku, rozhodli sme sa odsimulovať tento parameter v programe OptiSystem. Navrhnutá bola jednoduchá schéma, zobrazená na Obr. 2, pre skúmanie dostupnosti a chybovosti FSO systému vplyvom hmly. Parametre používané v simuláciách boli odvodené z Tabuľka 1. Tie sú zobrazené v Tabuľka 2. Vo všetkých meraniach bola nastavená fixná hodnota vzdialenosti a to 250 m, keďže naše FSO hlavice sú umiestnené v areáli TUKE práve v tejto vzdialenosti. FSO hlavica FlightStrata 155E pracuje s vlnovou dĺžkou 850 nm, maximálna prenosová rýchlosť je 155 Mb/s a maximálny vysielač výkon je 20W. Keďže prenášame na vzdialenosť maximálne 250 m, našou úlohou bolo zistiť, aký maximálny útlm nebude mať vplyv na kvalitu prenosu a akú chybovosť budú dosahovať jednotlivé hodnoty útlmu hmly.



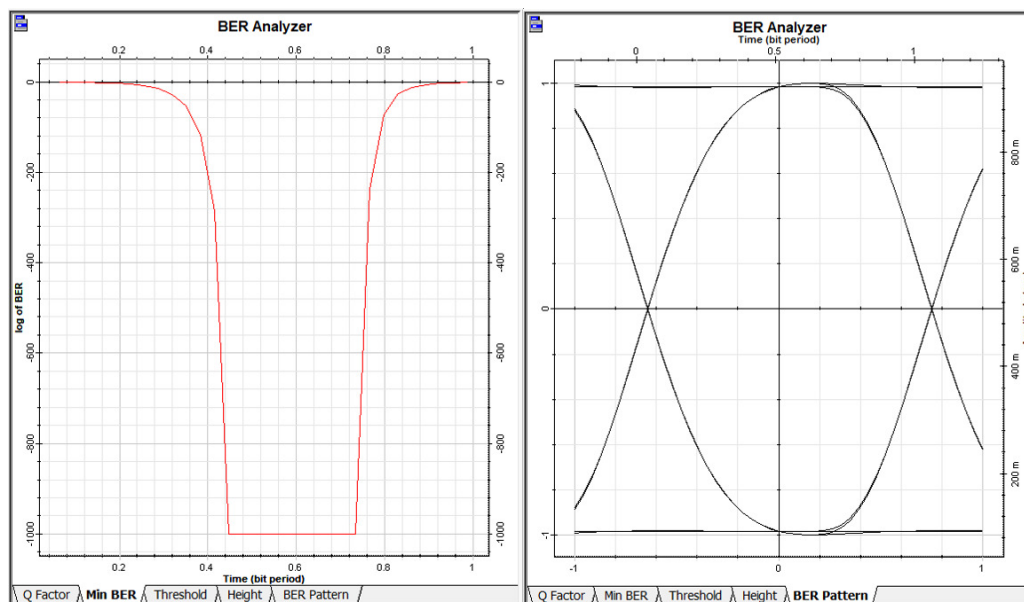
Obr. 2 Schéma zapojenia FSO systému v programe OptiSystem.

Tabuľka 2 Zvolené hodnoty útlmů hmly pre simuláciu v programe OptiSystem.

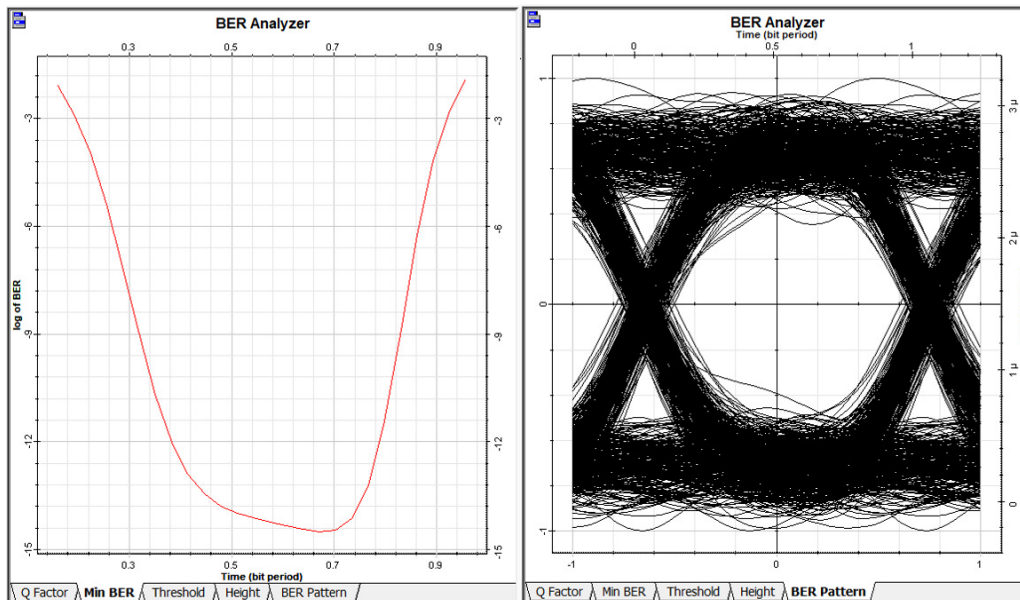
Hodnota útlmu [dB/km]	Q-faktor	BER
18,3	2275,5	0
29,8	2274,42	0
75	2245,16	0
150	952,264	0
200	75,1489	0
240	7,79897	$3,12045 \times 10^{-15}$
250	4,39226	$5,60853 \times 10^{-6}$
315	0	1

Ako je možné vidieť v Tabuľka 2, pre hodnoty útlmů do 200 dB/km je chybovosť (BER) 0 a hodnota kvality prenosu (Q-faktor) je veľmi priaznivá. Avšak, ak začneme postupne zvyšovať hodnotu útlmu, adekvátne k tomu začne klesať hodnota Q-faktor a BER. Pri našej vzdialenosti 250 m je maximálny možný útlm na hodnote 240 dB/km, kedy vieme povedať, že chybovosť pri tomto prenose je stále veľmi dobrá. Už pri hodnote útlmu 250 dB/km je možné vidieť, že Q-faktor má nízku hodnotu a chybovosť dosahuje veľmi malé hodnoty, preto pri tomto útlme môžeme povedať, že FSO linka nedokáže prenášať dostatočne dobre.

Obr. 3 zobrazuje hodnoty BER a diagram oka pre útlm 18,3 dB/km. Chybovosť dosahuje priam ideálne hodnoty, keďže je na úrovni 0, a rovnako tak na základe diagramu oka vieme povedať, že svetelný lúč prenášaný pri tomto útlme na vzdialenosť 250 m je málo zašumený. Ak sa postupne zvyšovala hodnota útlmu, klesala chybovosť a signál sa viac zašumoval. Na Obr. 4 je vidieť zmenu v chybovosti, kedy hodnota BER dosahovala úroveň $3,12045 \times 10^{-15}$. Z diagramu oka je zrejme, že pri útlme 240dB/km je prenášaný svetelný lúč značne zašumený a ovplyvnený práve týmto atmosférickým javom. Pri tejto hodnote útlmu vieme stále povedať, že nemá zásadný vplyv na prenos signálu, čo vyjadrujú aj dosiahnuté výsledky zapísané v Tabuľka 2.



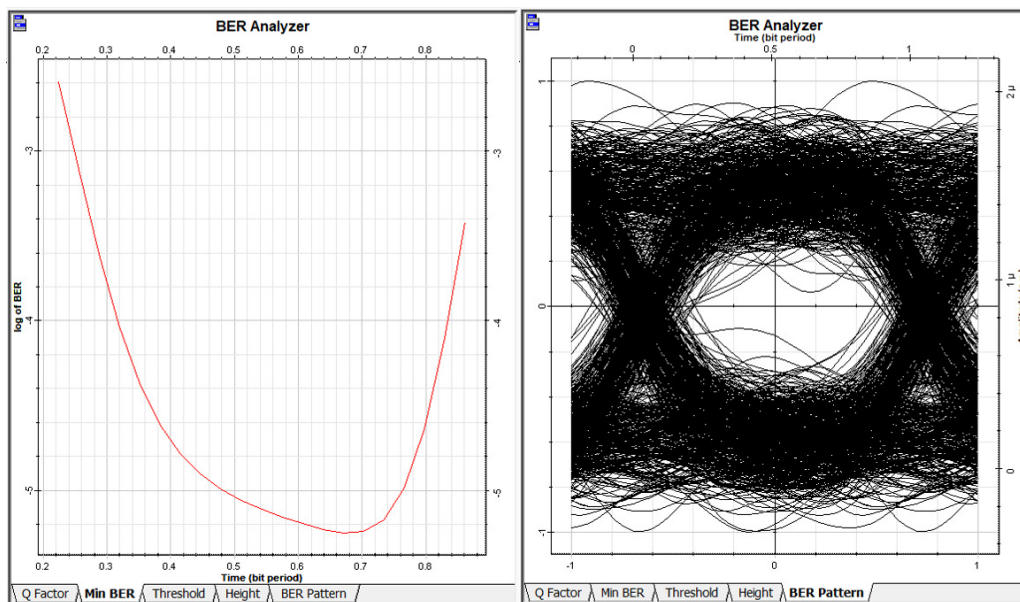
Obr. 3 Hodnota BER a diagram oka pre útlm 18,3 dB/km.



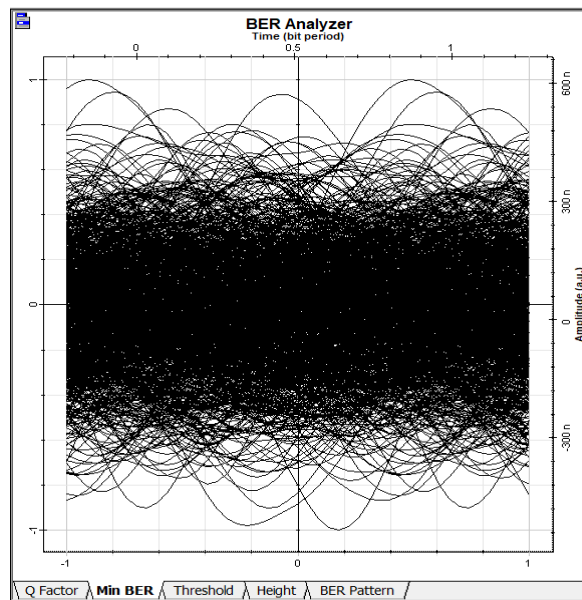
Obr. 4 Hodnota BER a diagram oka pre útlm 240 dB/km.

Pri pohľade na Obr. 5 je hneď vidieť výrazné zhoršenie kvality prenosu pri útlme 250 dB/km v porovnaní s predchádzajúcim výsledkom simulácie pri útlme na úrovni 240 dB/km. Chybovosť klesla oproti hodnote $3,12045 \times 10^{-15}$ len na úroveň $5,60853 \times 10^{-6}$, čo predstavuje pokles o 10^9 prijatých bitov. Z výsledkov v Tabuľka 2 vyplýva, že hodnota kvality prenosu je v porovnaní s predchádzajúcou hodnotou takmer o polovicu menšia, pričom útlm narástol len o 10 dB/km.

Najväčší vplyv na kvalitu prenosu a chybovosť BER mal útlm 315 dB/km. Pri tejto hodnote útlmu dosahovala kvalita prenosu hodnotu 0 a chybovosť bola na úrovni 1. Z Obr. 6 vieme s určitosťou povedať, že pri vzdialenosti 250 m a danej hodnote útlmu 315 dB/km je prenos pre náš FSO systém nerealizovateľný. Ak by sme uvažovali o tomto útlme, vzdialenosť prenosu medzi oboma hlavicami by bola rádovo v niekoľkých 10-tkách metrov.



Obr. 5 Hodnota BER a diagram oka pre útlm 250 dB/km.



Obr. 6 Diagram oka pre útlm 315 dB/km.

IV. ZÁVER

Táto práca bola zameraná na skúmanie dostupnosti a chybovosti FSO systému, na ktorý vplýval atmosférický jav hmly. Pri dosiahnutých výsledkoch vieme s určitou povedať, že pre nami zvolený systém s hlavicami FlightStrata 155E je maximálna hodnota útlmu na úrovni 240 dB/km, kde dosahujeme veľmi dobré výsledky. Zvyšovaním hodnoty útlmu úmerne klesá kvalita prenosu, ako aj chybovosť BER. Pre najväčšiu hodnotu útlmu (315 dB/km) už prenos nie je realizovateľný. Pre ďalšie skúmanie by bolo vhodné zobrať do úvahy aj ďalšie vplyvy atmosféry na prenosový kanál FSO systému a skúmať jednotlivé javy a ich vplyv na kvalitu prenosu a chybovosť.

POĎAKOVANIE

Táto práca bola podporená výskumným projektom FEI-2022-84 „Techniky spracovania údajov vo vysokorychlostných prenosových systémoch“.

LITERATÚRA

- [1] TechTarget, „What Is free-space optics,“ 2011, dostupné na: <https://www.techtarget.com/whatis/definition/free-space-optics-FSO>, [citované: 21.05.2022]
- [2] S. A. Al-Gailani, M. F. M. Salleh, a spol., „A Survey of Free Space Optics (FSO) Communication Systems, Links, and Networks,“ IEEE Access, 2021, roč. 9, 21 s., DOI: 10.1109/ACCESS.2020.3048049.
- [3] Z. Rahman, S. M. Zafaruddin, a spol., „Performance of Opportunistic Receiver Beam Selection in Multiaperture OWC Systems Over Foggy Channels,“ IEEE Systems Journal, 2020, roč. 14, zv. 3, DOI: 10.1109/JSYST.2019.2962580.
- [4] M. S. Khan, S. S. Muhammad, a spol., „Further results on fog modeling for terrestrial free-space optical links,“ Optical Engineering, 2012, roč. 51, zv. 3, 10 s., DOI: 10.1117/1.OE.51.3.031207.
- [5] A. Alatrash a E. Matida, „Characterization of Medication Velocity and Size Distribution from Pressurized Metered-Dose Inhalers by Phase Doppler Anemometry,“ Journal of Aerosol Medicine and Pulmonary Drug Delivery, 2016, roč. 29, zv. 0, str. 1 – 13, DOI: 10.1089/jamp.2015.1264.

Modelovanie dronov v kolaboratívnom virtuálnom prostredí

¹Branislav SOBOTA, ²Miriama MATTOVÁ, ³Lukáš JASENKA, ⁴Štefan KOREČKO

^{1,2,3,4}Katedra počítačov a Informatiky, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Letná 9, 04200 Košice, Slovenská republika

¹branislav.sobota@tuke.sk, ⁴miriama.mattova@tuke.sk

Abstrakt — Tento príspevok sa zaoberá modelovaním dronov v kolaboratívnom virtuálnom prostredí s využitím systému LIRKIS G-CVE. LIRKIS G-CVE je vyvíjaný na pracovisku autorov ako implementačná báza pre systémy kolaboratívnej virtuálnej reality. Príspevok obsahuje stručný úvod do problematiky dronov a rozšírenej reality (XR) s fokusáciou na vytvorenie virtuálneho prostredia pre simuláciu letu dronov. Riešenie, vzhľadom na použitie systému LIRKIS G-CVE využíva rámec A-frame, ktorý je jednoduchý pre vytváranie entít, najmä entity drona, avatara či nákladu a pridávanie komponentov pre určovanie správania a funkcionality týchto entít a ich prípadných fyzikálnych modelov. Pre dokonalé realistické prostredia v scéne je možné použiť viac modelov charakterizujúcich drona, avatara, nákladu alebo scén a máp. Opisom návrhu a implementáciou týchto entít sa zaoberá tento príspevok.

Kľúčové slová — dron, virtuálna realita, kolaboratívne virtuálne prostredia, simulácia

I. ÚVOD

V dnešnej dobe je veľmi často počuť o pojmoch, ako dron, bezpilotné stroje alebo o autonómnych leteckých prostriedkoch bez posádky. Drony sa označujú bezpilotnými lietajúcimi strojmi (angl. UAV, Unmanned Aerial Vehicle) [1]. Výmena informácií medzi používateľom a dronom prebieha spravidla pomocou ovládača, spolu tvoria určitý bezpilotný letecký systém. Ovládanie takéhoto typu zariadenia býva často doplnené aj určitou autonómiou, tým nazývame tieto drony autonómny. Výhody používania drona sú definované aj v [2]:

- spoľahlivosť a jednoduchosť konštrukcie,
- vysoká stabilita,
- ovládateľnosť v náročnom teréne,
- nízka vzletová hmotnosť,
- relatívne vyššia rýchlosť pohybu.

V súčasnosti existuje veľké množstvo rôznych typov dronov podľa ich konštrukcie a parametrov. Na základe zovšeobecnenia známych klasifikácií, taktických a technických charakteristík existujúcich lietadiel sa navrhuje klasifikovať ich podľa hlavných charakteristík ako:

- použitia v teréne,
- typu a tvaru krídel, ramien,
- režimu vzletu/pristátia,
- typu motorov,
- zdroj energie,
- hmotnosť.

Drony sa dnes spravidla konštruujú ako kvadrokoptéry či hexakoptéry. Označenie vyplýva z počtu vrtúľ resp. ramien. Príklady ďalších druhov dronov je možné nájsť napríklad v [3].

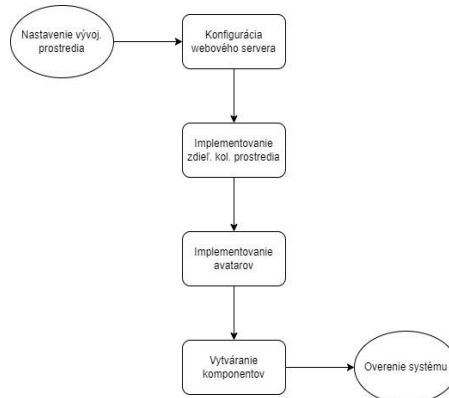
Tak ako mladá technológia dronov aj informačné technológie na čele s technológiami virtuálnej reality (VR) stále viac a viac prenikajú do života každého z nás. Narastá požiadavka neustále interaktívnejších a inteligentnejších (smart) riešení aj v oblasti dronov. Základnou aplikačnou platformou, ktorá by mohla úroveň interaktivity v oblasti dronov citeľne zvýšiť je rozšírená realita (XR) a v rámci nej virtuálna realita (VR) [4] a jej pridružené technológie ako augmentovaná realita (AR) alebo zmiešaná realita (MR) [5]. To zo sebou prináša aj neustále vylepšovanie rozhraní týchto systémov. Na pracovisku autorov príspevku (Laboratórium LIRKIS, KPI FEI TU Košice) sa vykonávajú práce, ktoré by mohli pomocou technológií XR v kontexte HCI a HCM-T [6] byť aplikovateľné v projekte Inteligentné operačné a spracovateľské systémy pre UAV aj v oblasti dronov. Tento koncept je zaujímavý aj z hľadiska možných

dlhodobých cieľov v kontexte rozvoja online tréningovania, zvýrazneného aj dopadmi predošlej pandemickej situácie.

Kolaboratívne virtuálne prostredia sú otvorené možnostiam rôznych tréningov, edukácii, simulácií alebo iných typov aktivít. Takáto platforma tak môže byť vynikajúcim základom pre tréning riadenia dronov nie len z pohľadu jedného drona ale aj hľadiska simultánnej kolaborácie viacerých dronov riadených viacerými operátormi. V [7] kolaboratívne systémy predstavujú špecifickú architektúru a technologické vybavenie pre možnosti vytvárania kolaboratívnych aktivít. Autori [8] vyzdvihujú aj potrebu fyzickej prítomnosti používateľov alebo ich virtuálnej podoby vo forme avatarov v kolaboratívnych prostrediach. Kolaboratívne virtuálne prostredia na tejto úrovni sa implementujú pomocou *virtuálnych jaskýň* alebo pomocou systémov založených na *virtuálno-realityných/dátových prílbách*. Vzhľadom na riešenie projektu a aj z hľadiska ceny aj dostupnosti sú VR/dátové prilby vhodným smerom, nakoľko je možné aj použitie VR prílb na báze smartfónu. Potom takto koncipované riešenie ako po stránke hardvérovej, tak po stránke programového vybavenia, s výhodou môže využívať systém LIRKIS G-CVE. Systém LIRKIS G-CVE [9], vyvíjaný na pracovisku autorov, navyše svojou implementačnou platformou (webové technológie) ďalej rozširuje aplikačnú bázu a znižuje cenu celkového výstupu. Samotný systém používa rámec A-frame a je zameraný na interakciu viacerých používateľov v jednej virtuálnej scéne využitím webového prehliadača. Riešenie, vzhľadom na použitie systému LIRKIS G-CVE, bude obsahovať kolaboratívne virtuálne prostredie a využité bude rámce A-frame, ktorý je jednoduchý pre vytváranie entít (ako entita drona a avatara) a pridávanie komponentov pre určovanie správania a funkcionality týchto entít. Pre dokonalé realistické prostredia v scéne bude použitých viac modelov charakterizujúcich drona alebo prostredie a mapy. Opisom návrhu a implementáciou práve týchto entít sa bude zaoberať tento príspevok.

II. NÁVRH FUNKCIONALITY A ENTÍT RIEŠENIA

Z pohľadu používateľa/ov by tento mal mať možnosť ovládania drona alebo iba sledovanie iných používateľov (spectator). Používateľ by mal mať možnosť si vybrať dron a špecifikovať jeho parametre, ako hmotnosť, veľkosť, senzory atď. Navyše by používateľ mal mať viacero možností ako môže ovládať dron, a to nie len klávesnicou. Mal by mať možnosť nahráť záznam o pohybe dronu a následne jeho spustenie ako autonómneho letu alebo jeho využitia ako tréningovej množiny pre umelú inteligenciu. V prostredí by sa mali nachádzať objekty, s ktorými môže dron hýbať, narážať, prípadne presúvať. Využitá by mala byť aj kooperácia, nie však vo forme zoskupenia dronov, ale systém, kde si môžu navzájom pomôcť, kde sa budú napr. striedať. Diagram postupnosti implementácie pre splnenie funkcionálnych cieľov je zobrazený na Obr. 1.



Obr. 1 Diagram postupnosti implementácie

Jednotlivé kroky implementácie teda pozostávajú najmä z:

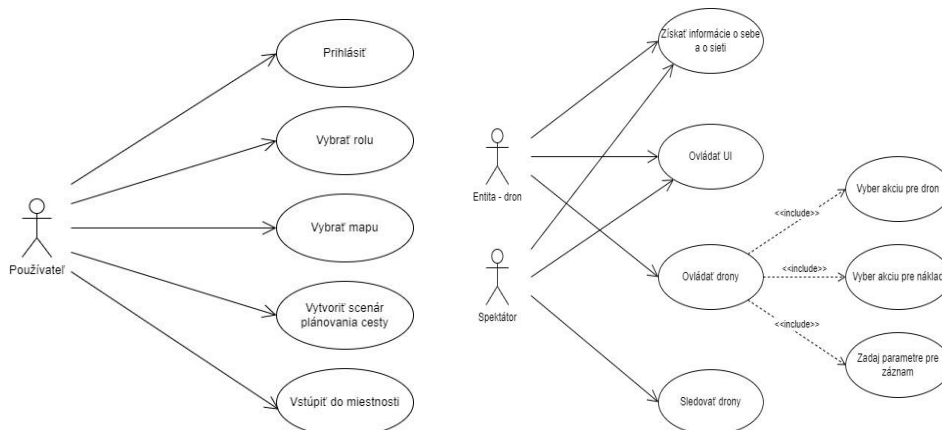
1. Nastavenie vývojového prostredia – predstavuje úvodnú konfiguráciu vývojového prostredia Glitch a vytvorenia projektu.
2. Konfigurácia webového servera – zabezpečenie komunikácie aplikácie a serveru.
3. Implementovanie zdieľaného kolaboratívneho prostredia – zabezpečenie vzájomnej interakcie medzi používateľmi a objektmi v scéne.
4. Implementovanie avatarov – vytvorenie avatarov, dronov, nákladov v podobe, vizuálnej aj funkčnej.
5. Vytváranie a implementovanie komponentov – spracovanie možností pre implementáciu interakcií akcií s dronom, nákladom, ovládaním a autonómnym letom.
6. Overenie systému – otestovanie pomocou scenárov rôzne interakcie medzi objektami a následné vyhodnotenie.

Z hľadiska rozhrania systému, vzhľadom na využitie LIRKIS G-CVE bude výhodné v tomto prípade využitie web-technológií. Rozhranie systému vyplýva zo schém uvedených na Obr. 2 a skladá sa z dvoch častí:

- rozhranie úvodnej stránky (Obr. 2 vľavo),
- rozhranie simulovania dronov (Obr. 2 vpravo).

Ďalej je potrebné navrhnuť systém implementácie, ako sa bude postupovať pri implementovaní servera, zdieľaného prostredia vrátane implementovania entít a komponentov. Vzhľadom na zložitosť a využitia technológií XR práve posledný problém je kľúčový a jemu sa budeme venovať v ďalšom. Ako teda vyplýva z doterajšieho, je potrebné navrhnuť hlavne tri entity (ako z hľadiska 3D virtuálneho modelu, tak aj z hľadiska parametrického a funkcionálneho):

- entita potrebná pre avatara,
- entita pre dron (dron môže tiež plniť funkciu avatara) a
- entita pre náklad.



Obr. 2 Schéma menu na najvyššom stupni (vľavo) a na úrovni simulovania dronov (vpravo)

A. Entita dron

Je reprezentovaná jednou triedou a niekoľkými podtriedami. Samotný dron obsahuje modelovú/vizuálnu stránku (Obr. 3), ktorá rieši jeho vzhľad, výzor a následná druhá stránka rieši napr. jeho fyzikálny model, špecifikácie či vlastnosti. Dronu je potrebné definovať:

- Fyzikálny model – správanie dronu podľa reálnych fyzikálnych zákonov (fyzikálny model), ako má telo reagovať na pôsobiace sily a kolízie v priestore, ďalej hmotnosť, rozmery tela, rozmery a rozmiestnenie vrtúľ atď.
- Stabilizáciu a výpočet rýchlostí motorov – PID regulátory, slúžia na stabilizáciu dronu počas letu, regulovanie otáčok motorov.

Akcie nad entitou dron:

- Vypnutie dronu,
- Pristátie dronu,
- Návrat dronu na miesto štartu,
- Vytvorenie záznamu,
- Pridanie bodu záujmu do záznamu,
- Spustenie záznamu, • Ovládanie pomocou klávesnice,
- Ovládanie pomocou ovládača.



Obr. 3 Príklad virtuálneho modelu drona

B. Entita avatar

Je reprezentovaná jednou triedou. Samotný avatar obsahuje vizuálnu stránku, ktorá rieši jeho vzhľad, výzor (Obr. 4) a druhá stránka rieši jeho špecifikácie, vlastnosti. Avatarovi je potrebné definovať:

- Fyzikálny model – správanie avatara podľa reálnych fyzikálnych zákonov, ako má avatar reagovať na pôsobiace sily a kolízie v priestore, ďalej hmotnosť a rozmery tela.



Obr. 4 Virtuálny model avatara

C. Entita náklad

Je reprezentovaná jednou triedou. Samotný náklad obsahuje vizuálnu stránku, ktorá rieši jeho vzhľad, výzor a druhá stránka rieši jeho špecifikácie, vlastnosti. Nákladu je potrebné definovať:

- Fyzikálny model – správanie nákladu podľa reálnych fyzikálnych zákonov, ako má náklad reagovať na pôsobiace sily a kolízie v priestore, ďalej hmotnosť a rozmery nákladu. Je potrebné riešiť zmenu fyzikálneho modelu drona aj nákladu v prípade, že je náklad naložený.

Akcie nad entitou náklad:

- Naloženie nákladu na dron,
- Vyloženie nákladu z dronu.

III. IMPLEMENTOVANIE ENTÍT

A. Implementácia kolaboratívneho virtuálneho prostredia

Pri tejto implementačnej časti bol využitý voľne dostupný externý aplikačný rámec *networked-aframe*(0.7.0). *Networked-aframe* slúži na vytváranie jednoduchých webových aplikácií určených pre zdieľanú interakciu viacerých používateľov využitím technológií *javascript* a *HTML* vo virtuálnom prostredí. V rámci dostupných funkcionalít má podporu pre online audio komunikáciu, je multiplatformový s podporou pre desktopové aj mobilné zariadenia a *HMD* systémy.

Úpravou konfigurácie na serveri bol zabezpečený spôsob sieťového prenosu údajov medzi zdieľanými *VR* scénami používateľov. Pre zabezpečenie zdieľania entity pomocou *networked-aframe* je potrebné vytvorenie šablóny zdieľaného objektu (*template*), samotnej zdieľanej entity a skriptu zdieľanej entity spolu s jej schémou.

B. Implementácia entity dron

Entita dronu predstavuje model, ktorému bola implementovaná fyzika pre pohyb a kolízie so statickými a dynamickými objektami. Pomocou rámca *A-frame* a jeho entitno-komponentového systému bol vytvorený model, ktorý sa skladá z hlavnej entity, ktorá charakterizuje jeho telo a obsahuje väčšinu komponentov, obsahuje komponent načítanie 3D modelu reprezentujúceho reálny model dronu. Ako pod-entity boli implementované dve entity pre rôzne pohľady na dron (obsahujú komponent kamera). Entita dronu je súčasťou zdieľaného prostredia a táto entita je zdieľaný objekt.

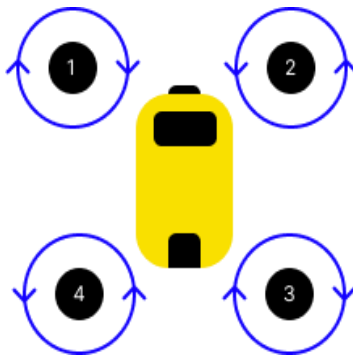
Pre implementovanie fyziky boli použité už implementované komponenty:

- *Dynamic-body* – komponent, ktorý je súčasťou balíčka *aframe-physics-system*. Balíček obsahuje komponenty, ktoré integrujú fyziku do rámca *A-frame*, podporuje ovládače *Cannon.js* a *Ammo.js*. Komponent definuje objektu entity hmotnosť a kolíziu obálku s inými objektami. Objekty sa môžu vplyvom fyziky odrážať, spomaľovať alebo voľne padať, ak je aktivovaná gravitácia.
- *Look-at* – komponent, ktorý je súčasťou balíčka *aframe-look-component*, komponent definuje správanie entity, aby sa otočila k inej entite alebo pozícii.

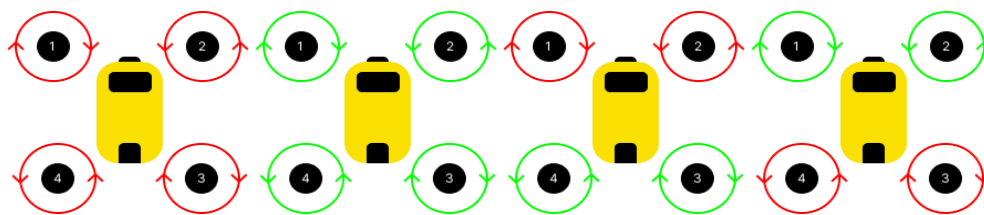
Pre implementovanie ovládania a stabilizácie boli vytvorené triedy:

- *Trieda pre ovládač drona* - sa stará o hlavnú časť. Spracováva dáta o fyzikálnej časti drona, údajov zo stabilizátora a ovládania drona. Údaje zo stabilizátora, výsledné hodnoty síl motorov, sú aplikované pomocou funkcie *applyLocalForce*, ktorá aplikuje silu na lokálnu

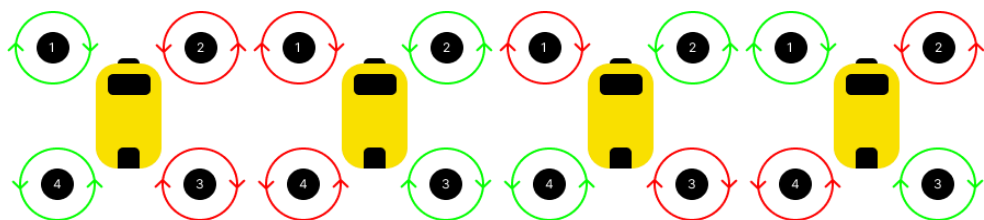
pozíciu drona, má dva parametre: *vektor sily* a *pozícia miesta pôsobenia sily*. Výsledné sily motorov však musia byť vyvážené, aby dron nestratil stabilitu. Hodnoty a porovnanie síl pre správanie sa drona je zobrazené na nasledujúcich obrázkoch (Obr. 5, Obr. 6, Obr. 7, kde zelená farba značí zvýšenie rýchlosti a červená zníženie).



Obr. 5 Hovering - súčet síl rovnaký gravitačnej sile



Obr. 6 Pohyb drona 1: nadol, dronu 2: nahor, dronu 3: dopredu, dronu 4: dozadu



Obr. 7 Pohyb drona 5: doľava, dronu 6: doprava; rotácia dronu 7: vpravo dronu 8: vľavo

- *Trieda pre stabilizáciu drona* – obsahuje implementovanie PID regulátora, ktorý je potrebný na stabilizáciu drona a výpočet rýchlosti motorov. Použitie niekoľkých PID regulátorov, práca s premennými, ako pozícia v priestore, rýchlosť pohybu, uhly natočenia a rýchlosť otáčania. Bolo vytvorených šesť PID regulátorov, pretože dron sa môže pohybovať v troch rôznych smeroch a nakláňať v troch rôznych uhloch. Vstupom je bod dosiahnutia, cieľový bod, a uhol natočenia v ypsilonovej osi (uhol medzi dronom a cieľovým bodom). Regulátory majú prístup k aktuálnej pozícii, rýchlosti pohybu, uhlom natočenia a rýchlosti otáčania. Regulujú tiež maximálne rýchlosti motorov a maximálne uhly natočenia. Následne tri regulátory vypočítajú hodnoty pre odchýlky od cieľového bodu, potom sa hodnoty použijú na výpočet uhlov náklonu pre každú os. Výpočet rýchlosti motora potrebné na dosiahnutie cieľového bodu, ktoré sú výstupom, získame nasledovne:

$$\begin{aligned}
 & \bullet \text{ Motor1} = \text{throttle} + \text{pitch} - \text{roll} + \text{yaw}; \\
 & \bullet \text{ Motor2} = \text{throttle} - \text{pitch} - \text{roll} - \text{yaw}; \\
 & \bullet \text{ Motor3} = \text{throttle} - \text{pitch} + \text{roll} + \text{yaw}; \\
 & \bullet \text{ Motor4} = \text{throttle} + \text{pitch} + \text{roll} - \text{yaw};
 \end{aligned}
 \tag{1}$$

kde hodnoty *pitch*, *roll*, *yaw* predstavujú výstupy pre otočenie v danej osi, výstupné hodnoty sú však ohraničené, aby nepresiahli maximálne rýchlosti motorov.

Ďalším komponentom bol komponent pre nastavenie farby entity.

C. Implementácia entity avatar

Entita avatara predstavuje model postavy, ktorý ovláda používateľ vo virtuálnej scéne. Pre pohyb a kolízie so statickými a dynamickými objektami je potrebné implementovať ovládanie pre pohyb a fyziku avatara. Pomocou rámca A-frame a jeho entitno-komponentového systému bola vytvorená postava avatara. Postava avatara sa skladá z hlavnej entity, ktorá charakterizuje jej telo a obsahuje väčšinu komponentov, ktoré prislúchajú entite avatar. Ďalšou pod-entitou je

entita, ktorá charakterizuje hlavu entity avatar, ktorá zabezpečuje rotáciu avatara a jeho pohľad (obsahuje komponent kamera). Obidve entity sú súčasťou zdieľaného prostredia, sú zdieľané objekty.

Pre ovládanie pohybu a fyzika avatara boli použité už implementované komponenty:

- *movement-controls* – komponent je súčasťou balíčka *A-Frame Extras*, ktorý ponúka prídavky a pomôcky pre rámce *A-frame* vo virtuálnej realite. Obsahuje komponenty pre ovládanie, načítanie modelov a viac. Komponent *movement-controls* zabezpečuje ovládanie avatara pomocou klávesnice, mobilných zariadení a ovládačov virtuálnej reality.
- *kinematic-body* - komponent je súčasťou balíčka *A-Frame Extras*, ktorý ponúka aj využitie ovládanie avatara v rámci fyziky a zabezpečuje kolízie avatara s ostatnými entitami, ktoré sú súčasťou virtuálnej scény. Tento komponent závisí na ďalšom balíčku *A-frame-physics-system*, ktorý obsahuje implementovanie podrobnejšej a rozsiahlejšej fyziky pre rámec *A-frame*.
- *look-controls* - Komponent, ktorý ponúka rámec *A-frame*. Komponent sa využíva väčšinou s komponentom kamery, aby sa zabezpečilo rotovanie entity, keď:
 - rotácia entitou pomocou myšky,
 - rotácia entitou pomocou mobilného zariadenia na dotykovom displeji potiahnutím,
 - rotáciou entitou pri otáčaní HMD.
- Komponent pre nastavenie farby entity

D. Implementácia entity náklad

Entita pre náklad predstavuje model, ktorému bola implementovaná fyzika pre určenie hmotnosti a reagovania síl, gravitácie. Pomocou rámca *A-frame* a jeho entitno-komponentového systému bola vytvorený model nákladu. Model nákladu sa skladá z jednej entity, ktorá má tvar kocky charakterizujúca jej telo. Entita je súčasťou zdieľaného prostredia, je zdieľaný objekt. Pre implementovanie fyziky boli použité už implementované komponenty:

- *Dynamic-body* – komponent, ktorý je súčasťou balíčka *A-frame-physics-system*. Balíček obsahuje komponenty, ktoré integrujú fyziku do rámca *A-frame*, podporuje ovládače *Cannon.js* a *Ammo.js*. Komponent definuje objektu entity hmotnosť, kolízie s inými objektami. Objekty nákladu sa môžu vplyvom fyziky odrážať, spomaľovať alebo voľne padať, ak je aktivovaná gravitácia.
- Komponent *constraint* slúži na spojenie telies pomocou pevných vzdialeností, je súčasťou balíka *A-frame-physics-system*.
- Komponent pre nastavenie farby entity

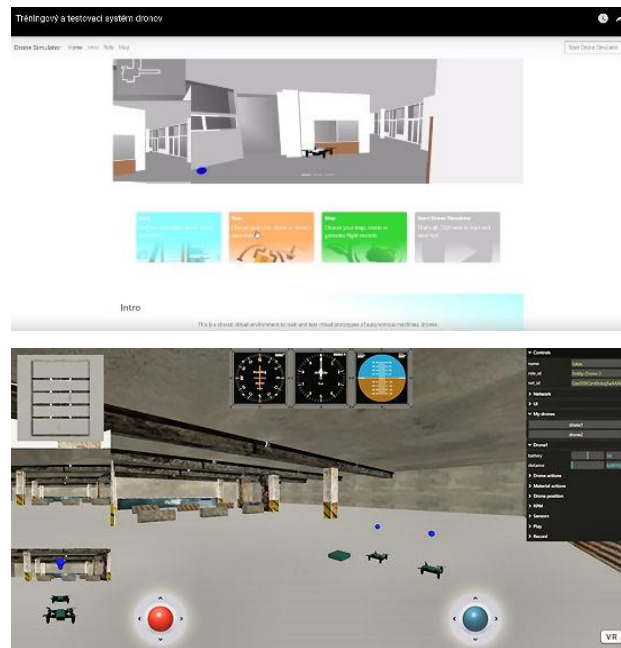
Pre implementovanie interakcie *dron a náklad* bola implementovaná trieda pre interakciu dron a náklad:

- *naloženie/vyloženie nákladu manuálne*: implementovaná funkcia, ktorá deteguje, či je náklad naložený. Ak nie je, skontroluje vzdialenosť medzi dronom a nákladom, v prípade vzdialenosti ideálnej pre reálne naloženie nákladu, náklad je pripojený k dronu. Ak náklad je naložený, potom je uvoľnený od drona z aktuálnej pozície drona.
- *naloženie/vyloženie nákladu automaticky*: implementovaná funkcia pracuje rovnakým spôsobom, ako predošlá. Rozdiel je v tom, že sa nezaobera vzdialenosťou, ale automaticky teleportuje náklad k dronu, alebo na miesto, kde sa náklad nachádzal na štarte.

IV. ZÁVER

Opisované riešenie obsahuje využitie kolaboratívnej virtuálnej reality s možnosťou zdieľania prostredia medzi používateľmi (operátormi) v oblasti tréningu a testovania dronov. Vytvorené riešenie (Obr. 8) ponúka webovú aplikáciu, kde úvodom je možné vybrať dron a typ mapy pre zobrazenie virtuálneho prostredia. Vo virtuálnom prostredí dronov používateľ (operátor) sa môže pohybovať pomocou avatara a ovládať jeden alebo viaceré drony.

V ďalšej práci by sme sa chceli zamerať na optimalizáciu výkonnosti pre zložité načítavanie 3D modelov a ich súčastí, optimalizáciou parametrov pre dron a využitia realistickejšieho fyzikálneho modelu. Veľkou výzvou bude aj prepojenie na služby umelej inteligencie, kde najmä implementovaný záznam letu drona riadeného operátorom môže slúžiť ako vhodná tréningová množina.



Obr. 8 Ukážka hlavného menu aplikácie (hore) a virtuálneho prostredia dronov (dolu)

POĎAKOVANIE

Táto publikácia bola podporená z operačného programu Integrovaná infraštruktúra v rámci projektu: Inteligentné operačné a spracovateľské systémy pre UAV, kód ITMS2014+: 313011V422 a spolufinancovaná Európskym fondom regionálneho rozvoja.

REFERENCIE

- [1] Ahmed, F.: Quadcopter : How it works and How it is made ? 2020. URL: <https://www.makermotules.com/quadcopter-how-it-works-and-how-it-is-made-2/>. (cit 05/2022)
- [2] Brown, J.: DRONEUSES: the awesome benefits of drone technology. 2022. URL: <https://www.mydronelab.com/blog/drone-uses.html>. (cit 05/2022)
- [3] Nagel, L.: Types of Drones and UAVs. 2021. URL: <https://www.tytorobotics.com/blogs/articles/types-of-drones>. (cit 05/2022)
- [4] Hudák, M.; Korečko, Š.; Sobota, B. : Enhancing Team Interaction and Crossplatform Access in Web-based Collaborative Virtual Environments, In: Proceedings of 2019 IEEE 15th International Scientific Conference on Informatics, IEEE, 2019, pp. 160-16.
- [5] Le Chénéchal, M., Duval, T., Gouranton, V., Royan, J., & Arnaldi, B. (2016, March). Vishnu: Virtual immersive support for helping users-an interaction paradigm for collaborative remote guiding in mixed reality. In 3DCVE 2016: International Workshop on Collaborative Virtual Environments (pp. 1-5). IEEE.
- [6] Sinha, G.; Shahi, R.; Shankar, M.: Human Computer Interaction. In: Emerging Trends in Engineering and Technology (ICETET), 2010 3rd International Conference on. IEEE, 2010. p. 1-4
- [7] Flavián, C., Ibáñez-Sánchez, S., & Orús, C. (2018). The impact of virtual, augmented and mixed reality technologies on the customer experience. Journal of Business Research.
- [8] Gonzalez-Franco, M., Hall, M., Hansen, D., Jones, K., Hannah, P., & Bermell-Garcia, P. (2015, March). Framework for remote collaborative interaction in virtual environments based on proximity. In: 3D User Interfaces (3DUI), 2015 IEEE Symposium on (pp. 153-154). IEEE.
- [9] Hudák, M.; Sobota, B.; Korečko Š. and Sivý M.: "Fully Immersive Web-Based Collaborative Virtual Environment for Upper Limb Rehabilitation Purposes," 2020 18th International Conference on Emerging eLearning Technologies and Applications (ICETA), 2020, pp. 194-199, doi: 10.1109/ICETA51985.2020.9379239.

Visualisation of highway rest stops data via mobile application

¹*Ema SUJOVÁ*, ²*Martin CHOVANEC*, ¹*Eva CHOVANCOVÁ*

¹Department of Computers and Informatics, Faculty of Electrical Engineering and Informatics Technical University of Košice, Slovak Republic

²Institute of Computer Technology, Technical University of Košice, Slovak Republic

¹ema.sujova@student.tuke.sk, ²martin.chovanec@tuke.sk, ¹eva.chovancova@tuke.sk

Abstract – The aim is to create a user-friendly mobile application that will primarily provide users with information about highway rest areas. We analyze similar existing solutions to our problems, types of mobile application development, technologies for cross-platform mobile application development and used front-end and back-end technologies, as we will use this knowledge in the implementation of our application. Furthermore, we describe the specific design and implementation of a mobile application solution.

Keywords – mobile application, flutter, python, cross-platform development, highway reststops

I. INTRODUCTION

Nowadays, we can do practically anything via a mobile phone - from a regular phone call to searching for the necessary information on the Internet to ordering food from a popular restaurant. In order to be able to make full use of our mobile phones, we need mobile applications. Mobile applications are used by people basically every day. There are a huge number of mobile applications for different purposes. Mobile applications have simply become an everyday part of us. We will be working on creating a mobile application *RestPoint* to help users on their highway journeys.

The goal is to create an application that will be intuitive to use, so that even a less technically proficient user can use it.

The application will provide users with information about services on individual rest areas located in Slovakia, provide them with the opportunity to register in the application and then save their favorite rest areas, provide them with assistance on the highway in any accident or event, from the application will be able to redirect directly to the main website of the National Highway Company, they will easily learn about the speed limits on the roads and much more.

II. EXISTING MOBILE TRAFFIC APPLICATIONS

In the analysis of our topic, we found three similar existing solutions to this problem, ie three mobile applications informing about the traffic situation and rest stops in Slovakia.

A. Highway assistance

The mobile application is intended mainly for calling the Highway Patrol via a call or SMS with the location of the person who needs help on the highway. The application also records current traffic restrictions from the central operator's workplace. It provides motorists with the opportunity to relax thanks to the choice of rest areas with their services. In addition to traffic information, the application also includes tourist attractions in Slovakia, which travelers can use on their way. The application contains the necessary contacts, current traffic news and charging information. [1]

B. About transport

The application is part of the National Traffic Information System of the Slovak Republic, which provides up-to-date information on transport on Slovak roads, such as traffic accidents, deadlines, restrictions and police patrols. The functionality of this application includes display of current traffic events in the area based on GPS location of the mobile phone, display of traffic situation on selected busy road sections along with images from traffic cameras, search for a railway stop and view the timetable of the connection and information about bike routes with their visualization on the map [2].

C. On road

The Austrian application offers up-to-date information on traffic and weather conditions on Austrian motorways and expressways. The functionality of the application includes the possibility of purchasing a digital motorway toll or a digital annual toll card, the user has a route planner for Austrian roads taking into account traffic information, an overview of all petrol stations and rest areas along motorways and expressways and an overview of all events and construction works along motorways and expressways. [3]

III. DEVELOPMENT OF MOBILE APPLICATIONS

Mobile application development is the process of creating software applications that run on a mobile device. Our goal is to create an application for as many users as possible, we decided to create it for iOS and Android platforms, which are the two most used mobile platforms in the world [4].

Native application development means creating a mobile application exclusively for one platform. The application is created using programming languages and tools that are specific to one platform.

Cross-platform applications can be built for two platforms and there is no need to write two different codes in two different programming languages. Therefore, we decided to follow the path of cross-platform development when creating a mobile application and we will use the Flutter framework with the Dart programming language.

IV. SOLUTION DESIGN

First of all, we want to display user-friendly resting places and data about them, as this is the main essence of our mobile application. These will be displayed on the map, which is the main page of the application. We also want to allow users to filter the rest areas based on some feature, such as fuel brand or whether the rest area has a shower or not. An integral part of the application is a button to call for help on the highway. We also want to give the user the opportunity to log in to the application through several forms of login, so that he can save his favorite rest areas and revisit them on the next trip on the highway.

In the process of designing the application, we first created the wireframes, which present the elements of content and the way of their distribution on individual subpages of the mobile application. The reason for creating wireframes is to design an information structure that will ensure that the future application will fulfill the expected purpose.

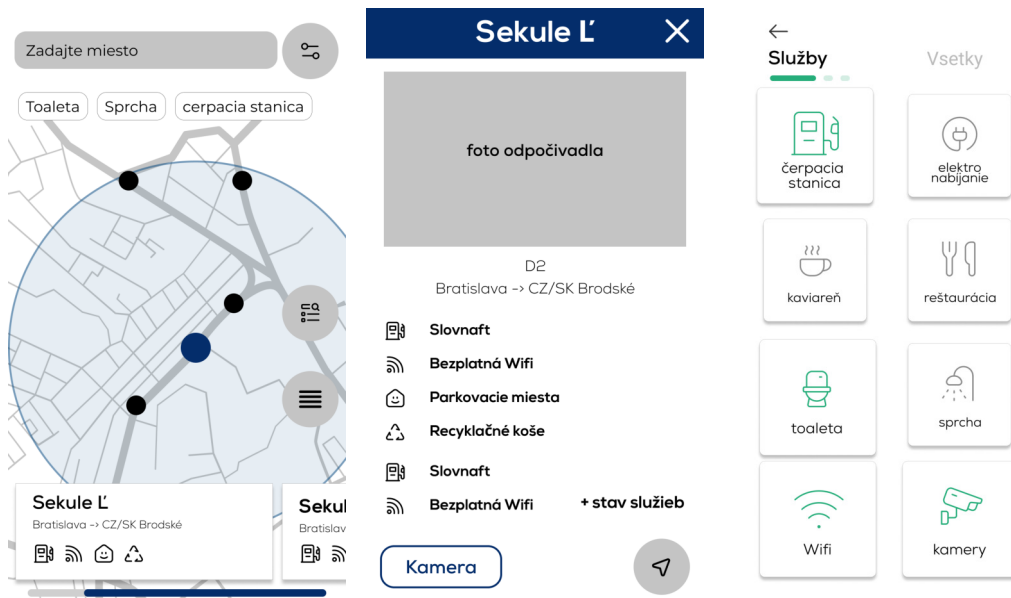


Fig. 1 Wireframes of our application

Based on the wireframes shown above, we can better imagine the final product. The next step in our design was to create a whole mobile application user interface based on wireframes.

Based on the wireframes, we created a prototype of the application that will be closest to what the final application will look like. Creating a quality application prototype will make it easier

for us to work in programming, as we will know how the individual components should interact and work.

For comparison, in the pictures below it is possible to see several pages of the application, which were displayed above as wireframes.

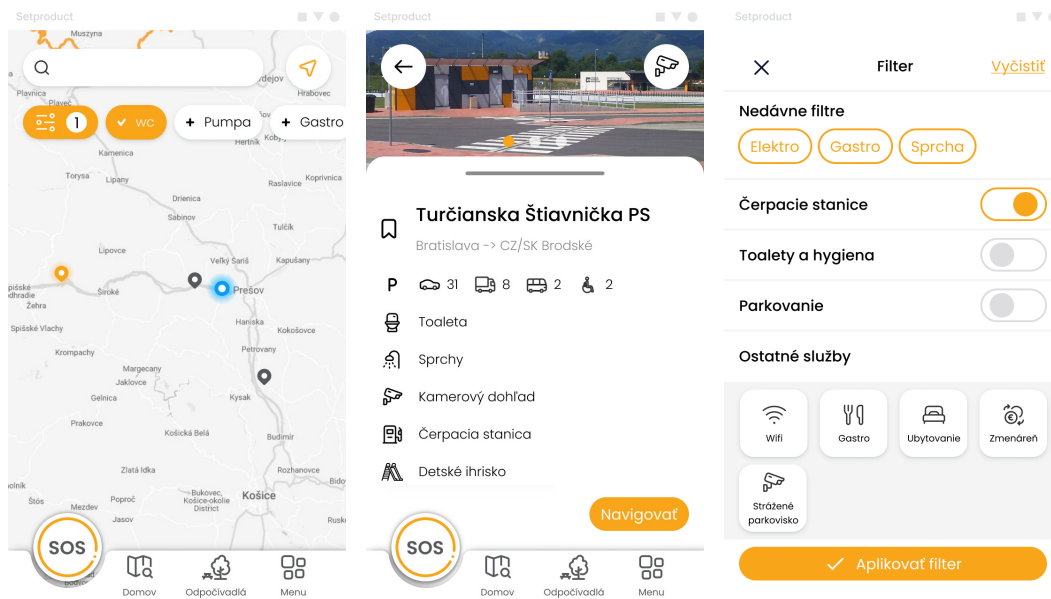


Fig. 2 UI design of our application

V. SOLUTION IMPLEMENTATION

When we talk about frontend and backend when developing an application, we divide the work for application development into two parts, as the creation of the frontend is independent of the backend and vice versa. This section will describe the technologies and procedures that were used in the implementation of the solution.

A. Frontend implementation

The frontend of the mobile application is the part with which the user directly interacts. Frontend development is mostly client-side development, which means that developers are involved in code analysis, application design and debugging, while ensuring a seamless user experience [5]. The frontend consists of everything a user sees when interacting with a web page or application, such as text colors and styles, photos, charts, tables, buttons, a navigation menu, and more.

Flutter is a high-performance framework based on the Dart language which provides high performance by rendering the user interface directly to the operating system canvas, rather than through the native framework [6].

One of the main components of the application is the display of rest areas on the map around the user. In order to do this, we need to view and track the user's location on the map while using the app. We get the user's current location using the Flutter add-on *location*, which handles location acquisition on Android and iOS systems. It also provides a callback when changing position.

Another essential part of the application is the filtering of rest areas based on one of the filters. It is a practical and usable functionality for users who want to view and then select the rest area they will visit, based on the features that the rest area has.

At the top of the home screen is a search box where the user can search for a city or rest area. If it searches for a city, the map zooms in on that city and shows at least the two nearest rest areas next to that city. If he searches for a rest stop, the user will be zoomed in on that rest stop.

At the bottom of the home screen is a larger SOS button, after which you can call for help from the highway patrol. Because the application monitors the user's location, it is only possible to call a highway patrol if the user is on the highway.

You can be redirected to several pages from the home screen. You can get to the list of rest areas by clicking on *Rest Areas* in the menu and, if you are logged in to the application, also to the list of saved rest areas.

After clicking on the menu button at the bottom bar, we will see the application menu, where you can find the greeting, *Sign In* button, travel charges, speed limits, highway patrol and you can send feedback directly to the highway patrol page. The app's home screen is shown below.



Fig. 3 Application homepage

The user can log in to the application in three ways, by logging in with a Google account, an Apple account, and an email and password. With the last option, it is necessary to first register, confirm the registration e-mail and then log in to the application with the given e-mail and password. By logging in, the user will be able to save their favorite rest areas in the application.

From the main screen, the user can view brief rest information. After clicking on the rest area, a large detail of the rest area will be expanded, where he will find all the information about it. Subsequently, it can be directly redirected to one of the navigation applications, which directly directs it to the rest area from its current position.

We will use Google Maps in the application. Using the Flutter add-on for Google Maps, we will add maps based on Google Maps data to our application. The add-on automatically handles access to Google Maps servers, map views, and responses to user gestures such as clicks and pulls. You can also add markers to the map depending on where the objects you want to appear on the map are located. We display the rest areas on the map through markers, each with its own exact location, which is the same as the geographical location of the rest area. We show three types of markers on the map: user location, rest area, and marked rest area.

B. Backend implementation

The backend is also known as the server side of the mobile application. Indicates any part of the software program that users do not see. It organizes, stores data and ensures that everything on the client side of the application works properly. Although the backend does not interact with users directly, it performs an irreplaceable task in the background by adding key features to the application. [7].

The backend includes activities such as writing APIs, creating libraries, and working with system components without user interfaces.

FastAPI is a modern, fast (high-performance), web framework for building APIs with Python 3.6+ based on standard Python type hints. FastAPI has very high performance, on par with NodeJS and Go (thanks to Starlette and Pydantic). [8] FastAPI is a Python framework and set of tools

that enables developers to use a REST interface to call commonly used functions to implement applications. It is accessed through a REST API to call common building blocks for an app.

We used Swagger to test and also document our API. Swagger is a REST API description language expressed using JSON [9]. Thanks to Swagger, an interactive API documentation is automatically created. So Swagger basically works by asking our API for a JSON, which contains a detailed description of our entire API.



Fig. 4 Swagger UI

Alembic is a database migration tool for use with the SQLAlchemy python database tool. Thus, SQLAlchemy is a set of Python SQL tools and an object-relational mapper that provides application developers with the full performance and flexibility of SQL [10]. We use the PostgreSQL database in the application and we use SQLAlchemy as the ORM over the given database.

We will use a token, specifically JWT, to secure the login API. JWT is used to share security information between two parties - the client and the server. Each JWT contains encrypted JSON objects, including a set of permissions, which distinguishes it from other web tokens.[11]

We needed to fill the database with data on rest areas. We obtained these from the file *read_excel.py*, from which we read data from tables with information about rest areas, and then we created the file *rest_stops.pkl*, which is created by a pickle module that allows serialization. objects to files on disk and deserialize them back to the program at runtime.

When storing rest areas in the database, we use their longitude and latitude so that we can display the given point where the rest area is located on the map. In order to obtain rest areas to some extent, we use an extension for the PostgreSQL object-relational database - PostGIS.

At the beginning, we define the point on the map around which we want to search for rest areas and then we choose the radius in which these rest areas should be located. We therefore use PostGIS to filter out rest areas around the city or user. We currently use this so that at least two landings must be displayed in the radius, otherwise the radius will increase and the search will continue. This is the optimal solution, as the user can just be far from the rest areas, and if the radius were statically set, he would not be able to find rest areas in his area.

In the application, we enable search in the list of places and rest areas, while the created records contain information about the name of the object, its type (city or rest area) and coordinates on the map. We have a list of rest areas and their location, but it was necessary to create the same list for Slovak cities. To create a list of cities, we obtained a list of all 141 Slovak cities from Wikipedia ¹. We extracted the names of Slovak cities from the table, which was located at the mentioned URL. After obtaining a list of Slovak cities, we used the REST API provided by the OpenStreetMap project to obtain their coordinates. We stored the obtained information about cities (name, latitude and longitude) in the pickle file *cities.pkl*, which we then used to create records in the data repository.

VI. EVALUATION

We tested the RestPoint mobile application on a group of 20 users to find out how our application is objectively perceived. We were interested in whether users are satisfied with the

¹https://sk.wikipedia.org/wiki/Zoznam_miest_na_Slovensku

form of logging in to the application that appealed to them most, whether they like the overall UI of the application, whether they found something that would hinder them in the application, and we asked about possible application improvements. Almost all users rated the application positively.

A. Application improvements

Every mobile application that has been created so far can be improved and thus attract more users. There is no difference in this application either. Thanks to testing, we also learned which improvements users would welcome. One is to add the number of currently active filters to the main screen of the application for a better user interface.

Another improvement that has been mentioned is the addition of opening hours for facilities and services located at the rest areas. Thanks to this improvement, users who want to use the services of restaurant facilities during their rest break, for example, would know if it will be open.

Adding the current number of free parking spaces on the rest areas is another functionality that users would welcome in the application.

Another possible improvement is to add current traffic information to the application so that users are informed about the current situation on highways. The addition of information on closed operations at motorway rest areas was also mentioned, thanks to which drivers would be able to plan their journeys more efficiently. Adding feedback directly to the rest area the user has visited is also a feature that could improve the quality of service at the rest areas, as it is now only possible to send feedback by e-mail to the National Highway Company.

VII. CONCLUSION

Mobile phones, especially smartphones, are an essential part of our lives today. Most of us use them every day and have an important role to play in every aspect of our lives. We mostly use various applications on smartphones, which has greatly affected and simplified our lives. From ordering food through online shopping to cleaning our house. Even when booking a taxi ride and road navigation, most of us use mobile applications. They have revolutionized the way we work, play, communicate and shop.

When we were designing the mobile application, we focused on creating user-friendly design, as the first reason for choosing any application from the user's point of view is its design. The application should be intuitive and easy to use for everyone.

We have implemented a multi-page mobile application in which users learn information about rest areas, they can search for a specific rest area on the map based on selected filters, they can call help for the highway and many other functionalities.

Testing the application on a group of users has brought us, among other things, possible improvements to the application, such as adding opening hours for restaurant facilities at rest areas and adding the number of currently available parking spaces at rest areas.

Application development is a continuous process and will continue even after the application is made available to all users, as it is in production that we get feedback from a larger number of users, ie what other functionalities would welcome in the application and reveal some bugs and shortcomings of the application. and testing overlooked.

ACKNOWLEDGMENT

This publication has been published with the support of the Ministry of Education, Science, Research and Sport of the Slovak Republic within project: Development and innovation of TUKE simulation environment in the field of technical sciences (004TUKE-2-1 / 2021)

This publication has been published with the support of the Operational Program Integrated Infrastructure within project: Research in the SANET Network and Possibilities of Its Further Use and Development (ITMS code: 313011W988), co-financed by the ERDF

REFERENCES

- [1] "Pomoc na diaľnici – aplikácie v službe google play," <https://play.google.com/store/apps/details?id=sk.nds.helpmotorway&hl=sk&gl=US>, 2022, [Online; navštívené 12-Apríl-2022].
- [2] "O doprave – aplikácie v službe google play," <https://play.google.com/store/apps/details?id=sk.telekom.odoprave&hl=sk&gl=US>, 2022, [Online; navštívené 12-Apríl-2022].
- [3] "Unterwegs – aplikácie v službe google play," <https://play.google.com/store/apps/details?id=at.asfinag.unterwegs&hl=sk&gl=AT>, 2022, [Online; navštívené 12-Apríl-2022].
- [4] "Cross-platform mobile frameworks used by global developers 2021 — statista," <https://www.statista.com/statistics/869224/worldwide-software-developer-working-hours/>, 2022, [Online; navštívené 12-Apríl-2022].
- [5] "Frontend definition," <https://techterms.com/definition/frontend>, 2020, [Online; navštívené 12-Apríl-2022].
- [6] "Flutter," https://www.tutorialspoint.com/flutter/flutter_tutorial.pdf, 2019, [Online; navštívené 12-Apríl-2022].

- [7] "Frontend vs backend - geeksforgeeks," <https://www.geeksforgeeks.org/frontend-vs-backend/>, 2021, [Online; navštívené 12-Apríl-2022].
- [8] "Fastapi," <https://fastapi.tiangolo.com/>, 2022, [Online; navštívené 12-Apríl-2022].
- [9] "Api documentation & design tools for teams — swagger," <https://swagger.io/>, 2011, [Online; navštívené 12-Apríl-2022].
- [10] "Sqlalchemy - the database toolkit for python," <https://www.sqlalchemy.org/>, 2021, [Online; navštívené 12-Apríl-2022].
- [11] "Json web tokens - jwt.io," <https://jwt.io/>, 2013, [Online; navštívené 12-Apríl-2022].

Analýza Výskytu Tsunami Pomocou Aplikácie na Podporu Rozhodovania

Martin MATTOŠ, Viera ANDERKOVÁ, František BABIČ

Katedra kybernetiky a umelej inteligencie, Fakulta elektrotechniky a informatiky,
Technická univerzita v Košiciach, Letná 9, 042 01 Košice, Slovenská republika

martin.mattos@student.tuke.sk, viera.anderkova@tuke.sk, frantisek.babic@tuke.sk

Abstrakt — Článok prezentuje vývoj interaktívnej webovej aplikácie na podporu rozhodovania v oblasti výskumu tsunami. Na tento účel boli použité historické dáta o tsunami od Národného úradu pre oceán a atmosféru. Pomocou jazyka R bola vybudovaná RShiny aplikácia, na základe ktorej dokáže odborník z oblasti tsunami analyzovať svoje dáta bez nutnosti použitia programovacieho jazyka. Aplikácia vo finálnej verzii obsahuje súbor funkcionalít, vďaka ktorým je možné dáta do aplikácie nahráť, vizualizovať vo vybraných grafoch, vykonať na nich základné aj pokročilé úpravy, pripraviť ich na proces modelovania a realizovať generovanie rozhodovacích modelov s možnosťou interpretácie výsledkov modelovania. Vývojárske testovanie aplikácie bolo realizované po každom dokončenom prototypu a v závere bola aplikácia otestovaná aj zástupcami cieľovej skupiny.

Kľúčové slová — podpora rozhodovania, CRISP-DM, tsunami

I. ÚVOD

Na svete existuje mnoho javov, ktoré majú negatívny dopad na našu planétu a tak aj na samotné ľudstvo. Do tejto problémovej skupiny sú zaradené aj živelné pohromy [1]. Medzi živelné pohromy sa radia udalosti, pri ktorých dôjde k nadmernému uvoľneniu energií vplyvom nežiadúcich prírodných síl. Tieto udalosti majú nepriaznivý dopad na život, majetok a zdravie, z čoho vyplýva, že medzi živelné pohromy patrí len také prírodné nebezpečenstvo, ktoré akýmkoľvek spôsobom vplyva na ľudskú spoločnosť. Charakteristickou vlastnosťou živelných katastrof je fakt, že nie je spôsobená priamym ľudským zásahom. Medzi základné živelné pohromy patria: zemetrasenia, tornáda, hurikány, lesné požiare, povodne, snehové kalamity, tropické cyklóny, extrémne teploty, zosuvy pôdy a aj nami skúmané tsunami. Najväčšou zaznamenanou vlnou v dejinách ľudstva bola udalosť v roku 1958, Lituya Bay, Alijaška [2]. Lituya Bay je malým zálivom (11 x 3 km) pri pohorí Fairweather. Zemetrasenie s magnitúdou 7,5 spôsobilo uvoľnenie približne 90 miliónov ton skál, ktoré po páde do zálivu vytvorili vlnu, ktorá vo svojom maxime dosahovala približne 520 metrov.

Za tsunami sa považuje vlna alebo za sebou idúca séria vln, ktoré sú vytvárané náhlým posunom stĺpca z vody [3]. Dôvodom vzniku takejto vlny/vln je hneď viacero. Najčastejšie sa jedná o seizmickú aktivitu – zemetrasenie pod hladinou [4]. V minulosti sa ľudia mylne domnievali, že strety litosferických dosiek pod hladinou oceánov sú jediným dôvodom vzniku vln. V skutočnosti môže tsunami vyvolať aj explózia, zosuv pôdy, dopad asteroidu alebo niektoré meteorologické javy (napr. silná búrka, vietor a podobne).

A. Analýza Súčasného Stav

Pri analýze existujúcich článkov alebo štúdií sme sa zamerali najmä na skúmané dáta, cieľ výskumu, použité metódy a algoritmy vo výskume udalostí tsunami.

Yanti a kol. [5] pracovali na včasnom varovaní pred tsunami použitím Bayesovho klasifikátora. Autori využili dáta z portálu InaTews o zemetraseniach a následných tsunami. Bayesovský klasifikátor bol naučený predpovedať potenciálne nebezpečenstvo tsunami na základe troch parametrov zemetrasenia: magnitúda, epicentrum a poloha. Experiment priniesol optimistický výsledok, keďže výsledná presnosť po použití tréningových dát bola 92,37% a f1-skóre malo hodnotu 0,98. Saengtattim a kol. [6] vytvorili prediktívnu analýzu poškodenia budov následkom tsunami. Cieľom bolo analyzovať a porovnať štyri atribúty maximálna (D_{max}) a kritická hĺbka toku (D_c), maximálna (V_{max}) a kritická rýchlosť toku (V_c) a následne nájsť optimálnu kombináciu

parametrov, ktoré by mohli byť najlepším ukazovateľom úrovne poškodenia budov. Pri modelovaní boli použité algoritmy: DTRA's: DT (rozhodovací strom), RF (náhodný les), GB (strom zosilnený gradientom) a algoritmus k-Means. Výsledky ukázali, že kombinácia atribútov D_c a V_{max} zaznamenala najvyššiu presnosť 89,56% pri klasifikácii metódami DT a RF, Naopak pre metódu GB bola zaznamenaná najvyššia presnosť pre dvojicu D_{max} a V_{max} a to 89,55%. Pughazhendhi a kol. [7] predikovali tsunami použitím kombinácie klasifikačného algoritmu strojového učenia a varovného systému pomocou IoT (Internet vecí). Pri skúmaní boli taktiež použité historické dáta o tsunami od Národného úradu pre oceán a atmosféru (z angl. National Oceanic and Atmospheric Administration, ďalej NOAA). Po spracovaní NOAA údajov mali k dispozícii spolu 995 platných vzoriek, z toho 249 bolo použitých na testovanie. Výsledná presnosť modelu RF bola 97,99% a skóre krížovej validácie pre daný model bolo 97%. Kusumah a kol. [8] sa venovali procesu určovania nebezpečenstva morskej vlny. Tréningové dáta obsahovali dva atribúty: výšku vlny, rýchlosť vlny a taktiež „označovací“ atribút, ktorý hovoril či bola daná vlna nebezpečná alebo nevytvárala ohrozenie. Ako testovacie dáta boli použité údaje z roku 2019 z Pangandarenského mora. Autori vo fáze modelovania využili metódu DT, konkrétne algoritmus C4.5, ktorý dosiahol presnosť v rozmedzí 98% - 100%.

Prostredníctvom analýzy súčasného stavu sme zistili, že existuje mnoho štúdií, ktoré sa zaoberajú problematikou tsunami. Autori sa venovali najmä: včasnému varovaniu pred prichádzajúcim nebezpečenstvom, klasifikácii „ničivosti“ rôznych udalostí v rámci niekoľkých rokov a odhadu sociálno-ekonomických dopadov tsunami. Identifikovali sme priestor na vývoj interaktívnej webovej aplikácie, ktorá by zoskupovala rôzne funkcie potrebné na vizualizáciu a analýzu dát pre výskumníkov v oblasti tsunami. Aplikácia by dodržiavala zásady a hierarchiu metodológie CRISP-DM a bez nutnosti poznania akéhokoľvek programovacieho jazyka by výskumníkom poskytla možnosť využívať prostriedky strojového učenia.

II. NÁVRH VLASTNÉHO RIEŠENIA

A. Vizia Riešenia

Našou víziou bolo teda vyvinúť jednoduchú, interaktívnu RShiny aplikáciu, na základe ktorej bude odborník z oblasti tsunami schopný analyzovať svoje dáta bez nutnosti použitia programovacieho jazyka. Aplikácia by obsahovala súbor funkcionalít, vďaka ktorým by bolo možné dáta do aplikácie nahráť, vizualizovať vo vybraných grafoch a mape sveta, vykonať na nich základné aj pokročilé úpravy, pripraviť ich na proces modelovania a v neposlednom rade generovať rozhodovacie modely s možnosťou interpretácie výsledkov modelovania. Základnou pomôckou pre bežného používateľa by boli informačné okná nachádzajúce sa pri názvoch sekcií. Používateľské rozhranie by sme vytvorili tak, aby bolo zrozumiteľné a jednoduché na používanie, s čitateľnou veľkosťou písma a v dobrom farebnom kontraste.

B. Cieľová Skupina a Požiadavky

Našu cieľovú skupinu tvorili odborníci na výskum tsunami (viď. Obr.1). Vo väčšine prípadov išlo o osoby, ktoré disponujú nanajvyššími základnými znalosťami z oblasti strojového učenia. Naša aplikácia by mala byť pomocným prostriedkom, vďaka ktorému bude aj laikom z tejto oblasti umožnená analýza dát na základe strojového učenia, bez nutnosti použitia programovacieho jazyka. Aplikácia však môže byť pomôckou aj pre vedcov, ktorí už majú so strojovým učením skúsenosti. V tomto prípade by mohla aplikácia uľahčiť prácu, ušetriť čas a tak zefektívniť proces analýzy dát o tsunami.

Následne sme vytvorili zoznam funkčných (viď. Tab.1) a nie-funkčných požiadaviek (viď. Tab.2) na vyvíjanú aplikáciu. Na stanovenie priority funkčných požiadaviek sme použili tzv. MoSCoW analýzu [9]. V prvej skupine sa nachádzali najdôležitejšie požiadavky, od ktorých závisela úspešnosť aplikácie. V druhej skupine boli požiadavky, ktoré by v aplikácii mali byť, no ak tam nebudú, nič sa nestane. Tretiu skupinu reprezentovali požiadavky, ktoré by bolo dobré mať v aplikácii, no ich realizácia by nemala pridanú hodnotu. V poslednej skupine boli požiadavky, na ktoré neboli alokované prostriedky a čas.

C. Informačná Architektúra Aplikácie

Keďže aplikácia mala dodržiavať hierarchiu metodológie CRISP-DM, bolo tomu prispôbené aj používateľské rozhranie. Práve karty v navigačnom menu prvej úrovne reprezentujú fázy spomínanej metodológie. V hlavnej časti každej z kariet sa postupne nachádzajú rôzne bočné panely či navigačné menu ďalších úrovní, vďaka čomu je zabezpečená vyššia prehľadnosť používateľského rozhrania (viď. Obr. 2). V týchto sekciách sa potom nachádzajú všetky pripravené funkcionality.



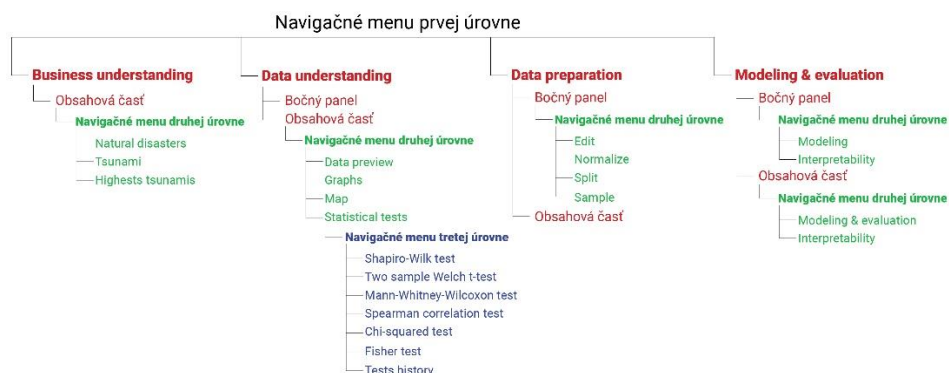
Obr. 1 Všeobecná persóna odborníkov v oblasti výskumu tsunami

Tab. 1 MoSCoW analýza funkčných požiadaviek

Možnosť nahratia vlastných dát do aplikácie	M
Možnosť zobrazenia dát vo forme tabuľky	
Možnosť vizualizácie dát pomocou grafov	
Možnosť vykonania základných úprav dát	
Možnosť vykonania prípravy dát na proces modelovania	
Možnosť generovania modelov pomocou rôznych algoritmov	
Možnosť porovnávania výsledkov generovaných modelov	
Možnosť zobrazenia štruktúry a sumarizácie dát	S
Možnosť vykonania zložitejších úprav dát	
Možnosť zisťovania závislosti/nezavislosti medzi premennými – štatistické testy	C
Možnosť editácie parametrov algoritmov v procese modelovania	
Možnosť interpretácie výsledkov metódou LIME	
Nahrané dáta nebudú ukladané na žiadne externé úložisko	W
Možnosť vložiť jeden záznam a pomocou trénovaného modelu vykonať predikciu	

Tab. 2 Zoznam nie-funkčných požiadaviek

Reakcia na voľbu používateľa prebehne maximálne do 3 sekúnd
Aplikácia nebude nikdy, bez zavinenia používateľa, neočakávane ukončená
V aplikácií budú všetky funkcie validované v dostatočnej miere
Text v celej aplikácii bude dobre čitateľný
Farebné prevedenie aplikácie bude mať vhodný kontrast
Aplikácia bude jednoduchá na používanie (z angl. user-friendly)
Bude dodržané pravidlo „troch kliknutí“, čo znamená, že ku každej funkcii sa bude možné dostať pomocou maximálne troch kliknutí
Pri každej sekcii sa bude nachádzať informačné okno
Aplikácia bude naprogramovaná v jazyku R s použitím balíčka shiny
Výstup aplikácia bude v offline podobe
Používateľské rozhranie aplikácie bude v anglickom jazyku



Obr. 2 Rozloženie používateľského rozhrania aplikácie

D. Implementácia riešenia

Pre proces návrhu, vývoja a s ním spojeného priebežného testovania sme vybrali historickú dátovú množinu tsunami od NOAA. Dáta sa skladali z približne 2 800 záznamov a 49 atribútov o udalostiach tsunami. Postupnými úpravami sme si vytvorili plnohodnotnú vzorku dát, ktorá obsahovala 445 záznamov a 16 atribútov a slúžila na priebežné testovanie.

V samotnej fáze vývoja aplikácie sme využívali prírastkový inkrementálny model [10]. Najprv sme zostavili zoznam konkrétnych funkcionalít, ktoré mala aplikácia obsahovať a pripravili plán práce. Potom sme prešli do procesu vývoja aplikácie, priebežného testovania a vylepšovania funkcií vo forme 4 prototypov. Predstavenie výslednej aplikácie cieľovej skupine a s ňou spojené záverečné testovanie sme uskutočnili až po kompletnom zhotovení aplikácie.

Základom prvého prototypu bolo vytvoriť rozloženie používateľského rozhrania - 4 karty v navigačnom menu prvej úrovne, ktoré korešpondovali s názvami fáz metodológie CRISP-DM a to „Business understanding“, „Data understanding“, „Data preparation“ a „Modeling & Evaluation“. Prvý prototyp obsahoval možnosť nahráť dáta do aplikácie, zobrazit' v podobe jednoduchej tabuľky, zmeniť dátový typ z číselného na kategorický, rozdeliť dáta na tréningovú a testovaciu množinu, zvolit' cieľovú premennú a generovať model z tréningových dát na základe algoritmu C5.0, kde bola výstupom sumarizácia modelu a jej graf.

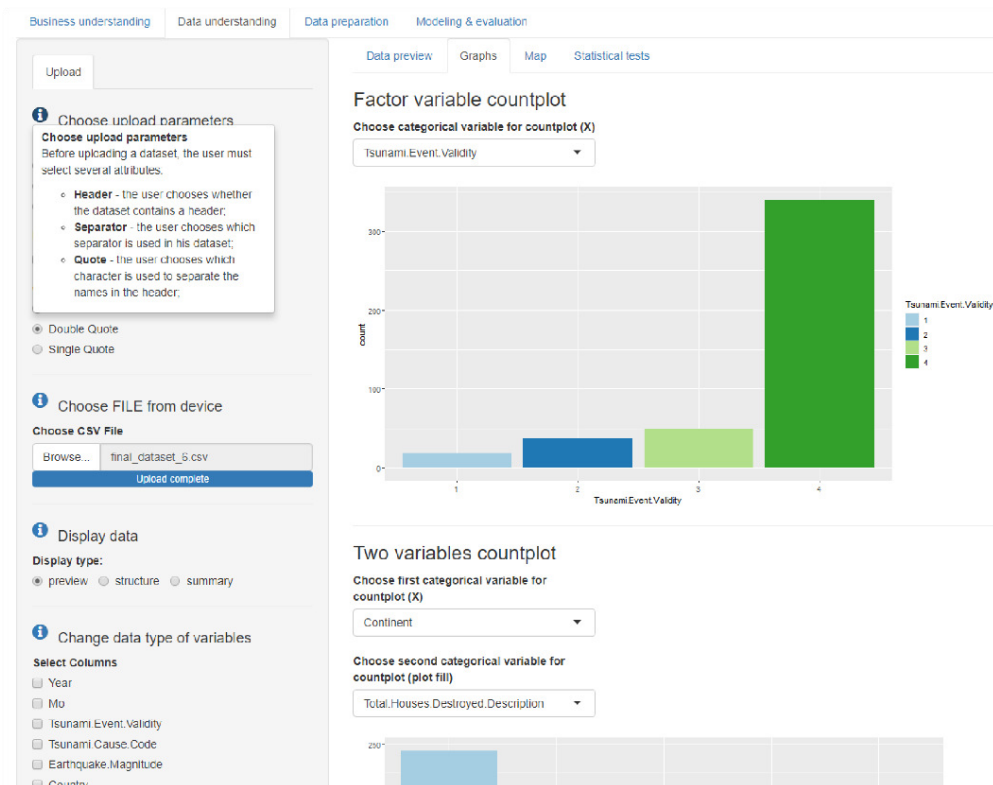
V rámci druhého prototypu aplikácie sme upravili karty „Data understanding“ o funkcionalitu zmeny dátového typu vybraného stĺpca a na karte „Modeling & Evaluation“ sme upravili funkciu delenia dát na podmnožiny a pridali sme informačné polia. Pridali sme aj nové rozšírenia ako vyplnenie karty „Business understanding“. Na kartu „Data understanding“ sme pridali vizualizáciu dát pomocou grafov (histogram pre kategorické a boxplot pre numerické dáta). Ďalej sme pridali funkciu zobrazenia všetkých nahraných dát v mape sveta. Na karte „Data preparation“ sme používateľovi poskytli možnosť vykonávať úpravu dát v podobe vymazania alebo premenovania vybraných stĺpcov, odstránenie riadkov s chýbajúcou hodnotou alebo riadkov s chýbajúcou hodnotou v špecifickom stĺpci. Na karte „Modeling & Evaluation“ sme k možnosti vybrať cieľovú premennú doplnili aj voľbu prediktorov, ktoré budú vstupovať do procesu generovania modelov.

V treťom prototypu bolo množstvo funkcií spracovaných do finálnej podoby, pridali sme však aj niekoľko nových funkcionalít takmer na každej karte navigačného menu prvej úrovne. Na kartu „Data understanding“ sme pridali možnosť vykonávať rôzne štatistické testy. Na kartu „Data preparation“ sme doplnili možnosť nahradiť chýbajúce hodnoty za priemer a strednú hodnotu prislúchajúceho stĺpca. Okrem toho sme pridali funkcie slúžiace na normalizáciu a diskretizáciu dát. Kartu „Modeling & Evaluation“ sme rozšírili o ďalších šesť algoritmov (caret, C5.0, RF, XGBoost, SVM s radiálnym kernelom, KNN a NB). Do fázy vyhodnocovania generovaných modelov sme pridali niekoľko metrik: AUC, F1-skóre, senzitivita a špecifickosť.

Štvrtým prototypom boli finálne doplnené a upravené všetky funkcie aplikácie. Hlavnou zmenou bolo pridanie možnosti editácie vzorkovania dát v tréningovej množine, doplnenie algoritmov o voľbu hodnôt parametrov a doplnenie interpretácie generovaných modelov pomocou metódy LIME. Konečná verzia karty „Business understanding“ obsahovala základné informácie o živelných pohromách a princípy tsunami. Na karte „Data understanding“ sa nachádzala možnosť nahrávania dát s výberom parametrov, zobrazenie dát v podobe náhľadu, štruktúry a sumarizácie, vizualizácia dát v štyroch rôznych grafoch, mapa sveta a štatistické testy. Karta „Data preparation“ obsahovala množstvo funkcií najmä v bočnom paneli: úpravy dát v podobe zmazania a premenovania stĺpcov, zmazania prvého riadku, všetkých riadkov s chýbajúcou hodnotou, náhradu chýbajúcich hodnôt za priemer alebo strednú hodnotu stĺpca, rôzne formy normalizácie a diskretizácie dát, voľbu cieľovej premennej a prediktorov, rozdelenie dát na tréningovú a testovaciu množinu vo zvolenom pomere, úpravu vzorkovania dát nad-/pod-vzorkovaním, náhľad dát v tréningovej množine a vizualizácia kategorických dát vo forme grafu.

E. Testovanie Aplikácie

Vývojárske testovanie jednotlivých funkcií prebiehalo súbežne s procesom vývoja aplikácie. Na základe týchto čiastkových testov sme odhaľovali chyby či príležitosti na vylepšenie funkcií pomocou pripravenej dátovej vzorky. Testovanie finálnej verzie prebiehalo aj pomocou dvoch zástupcov cieľovej skupiny z Univerzity Hradec Králové. Aplikácia bola hodnotená ako graficky prehľadná a dobre čitateľná. Všetky odporúčane návrhy na vylepšenie aplikácie sme zapracovali.



Obr. 3 Finálna verzia aplikácie

III. ZÁVER

Výskum tsunami je veľmi náročný proces, ktorému sa v súčasnosti venuje veľké množstvo organizácií a odborníkov po celom svete. Progres v tomto výskume a vývoj nových technológií s ním spojený je badateľný najmä vďaka výrazne klesajúcej krivke úmrtnosti následkom živelných pohrôm. Vytvorená aplikácia môže prispieť k lepšej orientácii a spracovaniu aktuálne dostupných dát aj pre používateľov s minimálnymi znalosťami z oblasti strojového učenia alebo dátovej analytiky. Do budúca je vhodné zväziť rozšírenie aplikácie o predikciu výskytu tsunami pomocou pravdepodobnostných metód a možnosť vyskladať semi-automatizovaný pracovný tok analytických krokov.

POĎAKOVANIE

Táto práca vznikla vďaka podpore Agentúry na podporu výskumu a vývoja na základe grantov č. APVV-20-0232 a APVV-17-0550 a Vedeckej grantovej agentúry MŠVVaŠ SR v rámci grantu č. VEGA 1/0685/2.

ZOZNAM REFERENCIÍ

- [1] S. W. A. Gunn, *Dictionary of Disaster Medicine and Humanitarian Relief*. Springer New York, 2012.
- [2] C. L. Mader and M. L. Gittings, "MODELING THE 1958 LITUYA BAY MEGA-TSUNAMI, II," 2002.
- [3] E. Bryant, *Tsunami*. Springer, 2009.
- [4] Y. X. Hu, S. C. Liu, and W. Dong, *Earthquake Engineering*. Taylor & Francis, 1996.
- [5] D. Y. Liliana and D. Priharsari, "Tsunami Early Warning Detection using Bayesian Classifier," in *2019 2nd International Conference of Computer and Informatics Engineering (IC2IE)*, 2019, pp. 44–48, doi: 10.1109/IC2IE47452.2019.8940823.
- [6] K. Saengtattim *et al.*, "Predictive Analysis of the Building Damage From the 2011 Great East Japan Tsunami Using Decision Tree Classification Related Algorithms," *IEEE Access*, vol. 9, pp. 31065–31077, 2021, doi: 10.1109/ACCESS.2021.3060114.
- [7] G. Pughazhendhi, A. Raja, P. Ramalingam, and D. K. Elumalai, "Earthosys---Tsunami Prediction and Warning System Using Machine Learning and IoT," in *Proceedings of International Conference on Computational Intelligence and Data Engineering*, 2019, pp. 103–113.
- [8] Y. Kusumah, B. Irawan, and C. Setianingsih, "Sea Wave Detection System using Web-Based Decision Tree Algorithm," in *2020 10th Electrical Power, Electronics, Communications, Controls and Informatics Seminar (EECCIS)*, 2020, pp. 231–236, doi: 10.1109/EECCIS49483.2020.9263444.
- [9] D. Haughey, "Moscow method," *Proj. Smart*, p. 2011, 2011.
- [10] G. Hinkel, *Implicit Incremental Model Analyses and Transformations*. KIT Scientific Publishing, 2021.

Windows Service for Preprocessing and Static Analysis of Application Logs of X-ray Devices

¹Eva MROČOVÁ, ²Branislav MADOŠ, ³Anton BALÁŽ

^{1,2,3}Department of Computers and Informatics, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Slovak Republic

¹eva.mrocova@student.tuke.sk, ²branislav.mados@tuke.sk, ³anton.balaz@tuke.sk

Abstract – The paper deals with the design and implementation of a Windows service for preprocessing and static analysis of X-ray devices application logs. X-ray device application logs are files that contain events recorded by software, such as warnings, information events and errors. The created Windows service *AILogManipulationService* meets the objective, which was to retrieve the log files of the X-ray devices of Siemens-Healthineers, analyse them and move them to the appropriate directories. Then create a model for AI log analysis through web application *Log Tool* and send notification emails to the responsible people, if the model is not created by the next time the service is run or if the current log file contains an error record. The automated and fully configurable *AILogManipulationService* service saves time and effort in separating log files into logs that contain an error message, and without an error message. For logs with no error message, models for further analysis are created via the web application *Log Tool*, owned by Siemens-Healthineers.

Keywords – Windows service, log file, SMTP client, static analysis of application logs, log separation, röntgen

I. INTRODUCTION

Nowadays, almost everyone has some experience with X-ray equipment. These devices are constantly evolving and improving their quality. Diseases can be detected earlier with the help of X-ray, which makes their treatment more effective. Radiography systems from Siemens Healthineers provide excellent imaging and workflow optimization with digital, analog and mobile X-ray systems. The software part of the radiography systems is the imaging system, which consists of several subsystems, e.g. the Unified Imaging System (UIS) and HarmonIC (Harmonized Image Chain). All software applications and systems produce log files. They are used as a source of information to understand the operation of computer systems and to monitor their health and stability. X-ray equipment application log files are files, which contain events recorded by the software, e.g. warnings, information events and errors. This paper discusses the implementation of a Windows service that is used for preprocessing and static analysis of UIS and Image Chain (IC) logs, where IC logs are created by the HarmonIC subsystem. The goal of the Windows service is to read the log files and, based on the result of the analysis, to separate them into the appropriate VALID and INVALID directories. Afterwards, the service sends notification emails to the responsible people, if a log file with an error message has been found. Finally, it automatically creates a model for AI (*Artificial intelligence*) analysis via the *Log Tool* web application.

II. ANALYSIS OF PRACTICAL IMPLEMENTATION OF WINDOWS SERVICES IN C#

Windows services are software programs, that are an essential part of the Microsoft Windows operating system [1]. They allow the user to create and manage long-running applications [2]. They can be started without user intervention and continue to run long after the user has logged out. These services run in the background and usually start automatically when the machine boots. They can also be paused and restarted. They do not display any user interface. These features make Windows services ideal to be used on a server or whenever long-term functionality is required, that does not interfere with other users running on the same system [3]. In Windows, the Management Manager manages all Windows Services processes (*Service Control Manager (SCM)*), which starts

at system boot. It is remote procedure call (RPC) server, so that programs service configuration and service management programs can manipulate services on remote machines [4]. SCM maintains a database of installed services and service drivers and provides a uniform and secure way to manage them. The database contains information about, how each driver service or service should be started. It also allows system administrators to customize the security requirements for each service, thereby controlling access to the service [5]. SCM displays a list of services and for each of them the Name, Description, Status (*Started*, *Suspended* or *Suspended*), Startup Type and Log On As. Each service can be stopped, paused, started, delayed or resumed as required.

Windows services can be easily created through the Visual Studio development environment or the .NET Framework SDK (Software Development Kit). Using the functions of the framework, it is possible to create, install, start and otherwise control the service behavior. When creating a Windows service, the Windows Service project template in Visual Studio can be used. The template will automatically do the groundwork - referencing the appropriate classes and namespaces, setting up base class inheritance for the service, and overriding a few methods that will be needed [7]. Windows service applications are class-based and inherit from the *System.ServiceProcess.ServiceBase* class. The main classes involved in creating services are [8]:

- 1) ***System.ServiceProcess.ServiceBase*** - the base class of the service that will exist as part of the Windows application of the service [9]. The user-created class inherits from this class, and the user also overrides its methods, which determine upon service state change. The *ServiceBase* class makes the *OnStart()*, *OnPause()*, *OnStop()*, *OnContinue()*, *OnShutdown()*, and *OnPowerEvent()* methods available, which can be overridden to add custom behavior. The properties of the *ServiceBase* class determine what methods can be called in the service. For example, when the *CanStop* property is set to *true*, it is allowed to call *OnStop* in the service. For the service to be useful, it is required to override at least *OnStart* and *OnStop*.
- 2) ***System.ServiceProcess.ServiceProcessInstaller* and *System.ServiceProcess.ServiceInstaller***
 - ***System.ServiceProcess.ServiceProcessInstaller*** – installs an executable file containing classes that extend the *ServiceBase* class. Installation utilities, such as *InstallUtil.exe*, call this class when installing the service application. This works for all services in the executable. This class is used by the installation helper to write registry values that are assigned to the installed service [10].
 - ***System.ServiceProcess.ServiceInstaller*** - works for the service it is associated with. The installer uses this class to write registry values that are assigned to a service to a subkey within the registry key HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\Services. A service is identified by its name in this subkey. The subkey also contains the name of the executable or .dll to which the service belongs [11]. The constructor of the *ServiceInstaller* class defaults the start type to *ServiceStartMode.Manual*, which specifies that the user starts the service. If the service needs to be started when the computer is turned on, the start type must be changed to *ServiceStartMode.Automatic*. It is required to set the *ServiceName* property of this class to match the name of the service to be installed.

For the installation helper to recognize the installer as valid, it is required to set the *RunInstallerAttribute* attribute to *true* in order for the program to call the installation methods.

III. UIS AND HARMONIC SUBSYSTEMS

A. Characteristics of the UIS subsystem

UIS is a project that will serve as a software platform of the RTG imaging system for the next 20+ years, enabling innovation, agile development and short time to market. It provides imaging system functionalities for all products used in medical clinics (radiology, urology, etc.) [12]. YSIO X.pree is the first product to bring UIS to the market. This product is designed to acquire images with excellent image quality. It brings a leap forward in radiography as it is equipped with *myExam Companion*. The UIS acquires images from the HarmonIC, which can be displayed on the screen and forwarded on to external systems located in the hospital. Using the DICOM interface, it performs data import or export. All activities are controlled by the primary user interface that is implemented in the UIS system. This primary user interface displays information about system settings, e.g. high voltage settings, as well as information on subsequent workflow steps. It also implements the *Administrator Portal*, which is used for settings and configuration at user and service level. In addition to the primary user interface, a touchscreen LCD display is provided and it is adjacent to the rack or tabletop assembly and serves as a secondary user interface with limited functionality compared to the primary user interface. This data display shows system parameters such as generator voltage settings, mechanical parameters such as table

or rack angles, and workflow related parameters.

The following logging facilities are available within the UIS system to analyze the behaviour of the display system in a retrospective manner [12].

B. Characteristics of the HarmonIC subsystem

HarmonIC calls detector API functions to initialize the acquisition, set acquisition parameters, etc. Once the acquisition is complete, the detector sends the two-dimensional image data to HarmonIC for further processing (brightness adjustment, contrast adjustment, cropping) [13]. Subsequently, HarmonIC sends the processed data via the C interface to the UIS subsystem. HarmonIC supports a control layer of various detectors and triggers. It also implements all image processing functions for X-ray systems during both - irradiation (real-time) and replay. It is therefore a major asset for the physical and medical aspects of image quality [14].

IV. X-RAY EQUIPMENT APPLICATION LOG FILES

Since the goal of the Windows service is to work with Image Chain (IC) and UIS log files that are obtained through tracing, it is reasonable to describe these logs in more detail.

A. UIS log file

The UIS log, shown in Fig. 1, is an XML file that consists of a combination of tag and character data. The tags give structure to the XML file, and the characters represent only the actual content. It has a hierarchical structure of elements, where each element can contain other elements and data of the represented object. In the UIS log, each message is made up of just one element, which contains its content element and additional elements with helpful information, such as

- **CreationTime** - the exact time when the message when it was written to the log file.
- **MachineName** - contains the name of the device.
- **Severity** - represents the message severity. Possible values: *Info*, *Warning*, *Error*.
- **Target** - defines whether the message is for the user or service. It can be *User* or *Service*.
- **MessageID** - unique message ID.
- **Message** - a message that carries information about the current activity or an error message.

```
<SDCMessageEvaluation Type="Message">
  <Item Name="CreationTime" Valuetype="xs:date">2021-07-01T00:24:16.332+02:00</Item>
  <Item Name="MachineName" Valuetype="xs:string">YSIO1</Item>
  <Item Name="Severity" Valuetype="xs:string">Info</Item>
  <Item Name="Target" Valuetype="xs:string">Developer</Item>
  <Item Name="MessageID" Valuetype="xs:string"></Item>
  <Item Name="Message" Valuetype="xs:string"><![CDATA[Set LockScreenNeeded flag to False]]></Item>
  <SDCMessageEvaluation Type="RuntimeParameters">
  </SDCMessageEvaluation>
  <Item Name="PID" Valuetype="xs:integer">12580</Item>
  <Item Name="ProcessName" Valuetype="xs:string">C:\UIS\bin\XP.UIS.SystemState.Container.exe</Item>
  <SDCMessageEvaluation Type="ExtendedProperties">
  | <Item Name="MsgCreateTime" Valuetype="xs:string"><![CDATA[2021-07-01T00:24:16.332+02:00]]></Item>
  </SDCMessageEvaluation>
</SDCMessageEvaluation>
<SDCMessageEvaluation Type="Message">
```

Fig. 1 Example UIS log file

B. IC log file

The IC log file is created right at HarmonIC startup and writing to the log file is terminated at shutdown. All workflows and other functions that harmonIC calls between its startup and subsequent shutdown are also written to the log file. The log, which is shown in Fig. 2, consists of 5 columns that contain the following information:

- **Date** - the date on which the message was written to the log file.
- **Time** - the time when the message was written to the log file.
- **ID thread** - the thread identification number is used to distinguish it from other threads. The ID has hexadecimal format.
- **Message type** - this part contains the *ERR* symbol if the operation was performed with an error, it contains the *ERR* symbol. On the other hand, if the operation was performed correctly, it is empty.
- **Message body** - it may contain one of the following:
 - 1) >>>> ENTERING - this symbol indicates the entry point into the function,
 - 2) <<<< LEAVING - this symbol indicates the exit point of the function currently being executed,
 - 3) further reports of the function that is currently being executed.

```

21.02. |11:32:37:824|100c| | | |
21.02. |11:32:38:826|100c| | | | |
21.02. |11:32:38:826|100c| | | | |
21.02. |11:32:38:826|100c| | | | |
21.02. |11:32:38:826|100c| | | | |
21.02. |11:32:39:826|100c| | | | |
21.02. |11:32:39:826|100c| | | | |
21.02. |11:32:39:826|100c| | | | |
21.02. |11:32:39:826|100c| | | | |
21.02. |11:32:40:826|100c| ERR|
(181)|
21.02. |11:32:40:826|100c| | | | |
21.02. |11:32:40:826|100c| | | | |
21.02. |11:32:40:826|100c| | | | |
21.02. |11:32:41:405|035c| | | | |
21.02. |11:32:41:405|035c| | | | |
21.02. |11:32:41:405|035c| | | | |
21.02. |11:32:41:405|035c| | | | |
21.02. |11:32:41:405|035c| | | | |
21.02. |11:32:41:405|035c| | | | |
21.02. |11:32:41:405|035c| | | | |
21.02. |11:32:41:405|035c| | | | |
21.02. |11:32:41:405|035c| | | | |
>>> ENTERING ipbasicutils::pingIPSendEcho - (132)
ipbasicutils::pingIPSendEcho - Call to IcmpSendEcho2 failed. (202)
ipbasicutils::pingIPSendEcho - Request timed out. (209)
<<< LEAVING ipbasicutils::pingIPSendEcho - ----- duration: 1.00243s (132)
>>> ENTERING ipbasicutils::pingIPSendEcho - (132)
ipbasicutils::pingIPSendEcho - Call to IcmpSendEcho2 failed. (202)
ipbasicutils::pingIPSendEcho - Request timed out. (209)
<<< LEAVING ipbasicutils::pingIPSendEcho - ----- duration: 1.00006s (132)
>>> ENTERING ipbasicutils::pingIPSendEcho - (132)
ipbasicutils::pingIPSendEcho - (Destination host was unreachable)
ipbasicutils::pingIPSendEcho - ICMP roundtrip time = 999 milliseconds (198)
<<< LEAVING ipbasicutils::pingIPSendEcho - ----- duration: 1.00001s (132)
>>> ENTERING ipbasicutils::pingIPSendEcho - (132)
>>> ENTERING hIC::Utils::ImageChainUtils::Utils::TraceVersion - (187)
Product Name: Healthcare harmoniC (279)
File: c:\harmoniC\bin\IC\hIC_CICudaFiltering.dll (280)
File Description: (281)
File Version: 0.1.2002.2101 (282)
Product Version: 0.1.2002.2101 (283)
Revision information: hIC_MG.Maintain.VE10.PackageAndBuild_MSBuild.Manual_20200221.1_C67299_MS
Build configuration: Release (285)

```

Fig. 2 An example of Image Chain log that contains an error symbol

V. WINDOWS SERVICE DESIGN

The first step of development is to obtain the requirements. It is one of the most crucial steps, as the entire product depends on this specification. Therefore, the functional and non-functional requirements of the Windows service are defined before the actual implementation. The functional requirements include: starting the Windows service, moving the data, loading the configuration file, checking the input, evaluating the data, generating the model for AI analysis, notifying the email client to send the email, and sending the email to the responsible people.

The non-functional requirements include: the Windows service must keep the processed information secure, adhere to GDPR conditions, be stable and extensible, and be configurable to e.g. add additional log file types.

Based on these requirements, a class diagram was designed to describe the structure of the system by showing the individual classes, their attributes, methods and relationships between objects. The graphical notation of the diagram is used to visualize the object-oriented system - Windows service [15]. The design also modelled behavioural diagrams, namely the use case diagram and the data flow diagram. The use case diagram shows the behaviour of the Windows service as seen by the user and describes the functionality of the system. The data flow diagram maps the flow of information. This section also designed the libraries/tools to implement the Windows service and to cover the different requirements that were used later in the implementation.

VI. IMPLEMENTATION

To create the Windows service, a template called *Windows Service* was used. The template is provided by the Microsoft Visual Studio environment. It automatically sets the namespace (*namespace*), inheriting from the base class *ServiceBase*, which provides methods such as *OnStart()* and *OnStop()* to represent the states that the service is in. This implemented template forms the basis of the work, to which other necessary classes were gradually added.

A. Creating and renaming the service

The template contains a class with the default name *Service1*. After renaming this to *AllLogManipulationService*, it was required to change the value of *ServiceName* in the *Properties* window on the [Design] tab to the name of the renamed class service. This class inherits from the base class *ServiceBase* and contains methods, such as *OnStart()*, *OnStop()*, *OnTimer()*, the enum data type *ServiceState*, and the structure *ServiceStatus*.

B. Adding logging

The template offers the possibility to write application events to a log file (*Event log*). Using the *AutoLog* property, it was possible to specify whether we want to use this feature. By default, this feature is enabled, but when creating Windows service, this feature was turned off, as the Windows service could implement a custom log file, using the *EventLog* class. To implement the log file, it was necessary to modify the existing constructor and import *System.Diagnostics*, which provides classes that allow interaction with system processes and log files. This log file stores information and exceptions of application events, e.g. when a service is started, the text defined in *OnStart()* method. The log file that stores information about the running service can be opened through the Event Viewer application.

C. Defining service behaviour

The implementation of the Windows service followed the proposed class diagram. The main class of the *AILogManipulationService* has defined methods *OnStart()* and *OnStop()*, which tell what state the service is in and how it should behave in that state.

- **Method *OnStart()*** - is called as soon as the service is started. At the very beginning, the service state is updated to *ServiceState.SERVICE_RUNNING*. Since the service is designed as a long-running application that polls or monitors the system, the method must return to the operating system after the service operation has started to avoid system deadlock. Therefore, it was necessary to implement cyclic polling where a timer is set. As soon as the timer elapses, the *OnTimer()* method is triggered at regular intervals by the *Timer.Elapsed* event.
- **Method *OnTimer()*** - started by default every hour, which can be changed in the configuration file. It contains almost all the functionality of the service and calls most of the implemented classes. This method calls the *Manipulation* class, which checks and moves log files to the *VALID* and *INVALID* directories. It calls a class to create a configuration file if it does not already exist. The configuration file, which is in XML format, contains the URL to create the model, the names and email addresses of the responsible people, the error symbol for the *UIS* and *IC* log files, the path to the log files, the data for the creation of the *SMTP* client and the start interval. If the method for checking log files of class *Manipulation* returns a list containing some log files, the *EmailClient* class is called. It contains methods that provide sending of notification emails. These methods are using *System.Net.Mail* and *System.Net*, which are required for the *SMTP* protocol. After this check, a model for AI analysis is created by a class that contains methods that prepare log files that do not contain an error record into the required form. A request to the server is sent to upload and to train the data. There is a possibility, that the creation of model will take several hours. Therefore, it was necessary to implement a check to see if the model has been created. If it is not created by the next time the method *OnTimer()* is run, an email is sent to the responsible people that informs them that the model could not be created. The interval of how often the *OnTimer()* method will be called is defined in the configuration file as an integer in hours. Given that, the interval for the timer is specified in milliseconds, unit conversion is being used.
- **Method *OnStop()*** - specifies the actions to be performed when the service receives a stop order. When the *SCM* receives this stop command, first it checks the value of *CanStop* to see if the service itself will receive a stop command. If *CanStop* is *true*, the service receives the stop command and the *OnStop()* method is called.

D. Adding installers to the service

User must install the service before he/she can start it. The installation will register the service in the *SCM*. To add the installers, it was necessary to view the [Design] main class of this service and select *Add Installer*, and afterwards add text to the properties e.g. *ServiceName*, *Description* and set *StartType* to *Automatic*.

E. Setting the service status

The service reports its status to the *SCM* so that the user can see if the service is working properly. The class, which inherits from *ServiceBase*, has a default set of states, including *SERVICE_STOPPED*, *SERVICE_PAUSED*, and *SERVICE_RUNNING*. If the service would take longer to start, the *SERVICE* status reporting *_START_PENDING* and also the *SERVICE_STOP_PENDING* if the service stopped entirely. The implementation consisted of using *System.Runtime.InteropServices*, declarations *ServiceState* values and adding the *SetServiceStatus* function.

F. Setup project

To run the implemented Windows service *AILogManipulationService* on another computer, it was necessary to create an installation by adding a new project called *Setup Project*. It was required to add a configuration file to this project, set the properties of the setup file, add the output of the project and compile it.

VII. BENEFITS

The created Windows service saves developer effort and time, e.g. when separating log files. Windows service is fully automated once started, which means that if you do not need to change the application configuration using a configuration file, there is no user intervention required. The service is also fully configurable, which, to some extent, allows the user to modify

its functionality. It includes information about its current status visible via the Event Viewer. Furthermore, it is adaptable to other types of log files. Therefore, the service can handle other types of log files but that requires some functional parts of the service to be changed. If *Log Tool* starts to support other log file types, then the Windows functionality of the service can be adapted as well. In order to run the created Windows service *AILogManipulationService* on other computers, an installation is provided, which, in addition to the Windows service, also contains the aforementioned configuration file with a path that can be specified by user during the installation.

VIII. CONCLUSION

This paper focused on the Windows service used to preprocess and do a static analysis of application logs of Siemens-Healthineers' X-ray equipment. To simplify the creation of the Windows service *AILogManipulationService*, the Windows service template offered by the Visual Studio environment was used. Functionality during its on and off times was also implemented for the correct functioning. The service automatically reads the necessary data from the configuration file that the service creates if it is not created. This file is used to efficiently edit individual *AILogManipulationService* settings. It is possible to modify the following data - path, where log files are stored on disk, the names and email addresses of the responsible persons, the error symbol of the UIS and IC logs, the data for the creation of the SMTP client, the startup interval, and the URL for the creation of the AI model. As part of the log file separation, they are automatically divided into two directories - VALID and INVALID. The INVALID directory contains only log files with error messages and the VALID directory contains logs without them. The service sends a request to upload, train the data and request a confirmation from the *Log Tool* web application that the model has been successfully created, using an HTTP client. The SMTP client sends notification emails to the responsible people if the service finds a log file that contains an error message, or if it failed to create a model in the *Log Tool* web application. The service aims to retrieve these log files, parse them and move them, either to the INVALID directory or VALID. It then creates a model for AI log analysis and sends notification emails in if the model is not created by the next time the service is run, or if the current log file contains an error record. The INVALID directory contains log files that contain the error record and therefore the error symbol, and the VALID directory contains the log files without an error symbol. Based on the testing performed, the functional and non-functional requirements that were proposed for this system have been met. Some of the requirements were, e.g., evaluating and moving data, sending notification email to responsible persons, creating AI model, loading configuration file, etc.

ACKNOWLEDGMENTS

This publication has been published with the support of the Operational Program Integrated Infrastructure within project: Research in the SANET Network and Possibilities of Its Further Use and Development (ITMS code: 313011W988), co-financed by the ERDF. This research was supported by KEGA 002TUKE-4/2021 Implementation of Modern Methods and Education Forms in the Area of Cybersecurity Towards Requirements of Labour Market.

REFERENCES

- [1] F. Pathan. (2020) Create a windows service in c#. [Online]. Available: <https://www.c-sharpcorner.com/article/create-windows-services-in-c-sharp/>
- [2] Microsoft. Introduction to windows service applications. [Online]. Available: <https://docs.microsoft.com/en-us/dotnet/framework/windows-services/introduction-to-windows-service-applications>
- [3] A. Altvater. (2017) What are windows services? how windows services work, examples, tutorials and more. [Online]. Available: <https://stackify.com/what-are-windows-services/>
- [4] Microsoft. Service control manager. [Online]. Available: <https://docs.microsoft.com/en-us/windows/win32/services/service-control-manager>
- [5] ——. About services. [Online]. Available: <https://docs.microsoft.com/sk-sk/windows/win32/services/about-services>
- [6] ——. (2017) Develop windows service apps. [Online]. Available: <https://docs.microsoft.com/en-us/dotnet/framework/windows-services/>
- [7] ——. (2017) How to: Create windows services. [Online]. Available: <https://docs.microsoft.com/en-us/dotnet/framework/windows-services/how-to-create-windows-services>
- [8] ——. (2017) Service application programming architecture. [Online]. Available: <https://docs.microsoft.com/en-us/dotnet/framework/windows-services/service-application-programming-architecture>
- [9] ——. Servicebase class. [Online]. Available: <https://docs.microsoft.com/en-us/dotnet/api/system.serviceprocess.servicebase?view=dotnet-plat-ext-5.0>
- [10] ——. Serviceprocessinstaller class. [Online]. Available: <https://docs.microsoft.com/en-us/dotnet/api/system.serviceprocess.serviceprocessinstaller?view=netframework-4.8>
- [11] ——. Serviceinstaller class. [Online]. Available: <https://docs.microsoft.com/en-us/dotnet/api/system.serviceprocess.serviceinstaller?view=netframework-4.8>
- [12] S. Healthineers, "Uis software architecture specification," unpublished, 2021.
- [13] —, "Sharpfn architectural view," unpublished, 2019.
- [14] —, "Image chain," unpublished, 2015.
- [15] V. Paradigm. Uml class diagram tutorial. [Online]. Available: <https://www.visual-paradigm.com/guide/uml-unified-modeling-language/uml-class-diagram-tutorial/>

Overenie vlastností logistického zobrazenia pomocou odhadu Ljapunovových exponentov

*Jakub ORAVEC, Ľuboš OVSENÍK, Maroš LAPČÁK,
Norbert ZDRAVECKÝ, Samuel ANDREJČÍK*

Katedra elektroniky a multimediálnych telekomunikácií,
Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach

{jakub.oravec, lubos.ovsenik, maros.lapcak, norbert.zdravecky,
samuel.andrejcik}@tuke.sk

Abstrakt – Tento príspevok sa venuje analýze logistického zobrazenia, ktoré patrí v oblasti obrazových šifrovacích algoritmov k najčastejšie používaným jednorozmerným chaotickým zobrazeniam. Vlastnosti tohto zobrazenia sa často kvantifikujú pomocou Ljapunovových exponentov. Táto práca sa snaží poukázať na rôzne premenné pri odhade hodnôt týchto exponentov, ako napr. konečný počet vzoriek pri odhade a konečná presnosť zápisu desatinných čísel. V stručnosti budú analyzované aj maximálna a minimálna hodnota odhadovaných Ljapunovových exponentov a ilustruje sa ich praktický význam.

Keľúčovú slová – Ljapunovove exponenty, logistické zobrazenie, obrazové šifrovacie algoritmy

I. ÚVOD

Dedikované obrazové šifrovacie algoritmy vznikli z dôvodu efektívnejšieho spracovania obrazových dát [1]. Preto je ich architektúra do istej miery odlišná od architektúry konvenčných šifrovacích algoritmov ako napr. Advanced Encryption Standard (AES) [2]. Obrazové šifrovacie algoritmy sa skladajú z blokov, ktoré môžu jednoducho spracovať vysoko korelované bloky dát s veľkým počtom vzoriek. Ku takýmto blokom patrí napr. preusporiadanie riadkov, resp. stĺpcov obrazov pomocou cyklických posunov, resp. kombinácia s pseudo-náhodnou (PN) postupnosťou pomocou funkcie exkluzívneho súčtu XOR (z angl. eXclusive OR). Aj v prvom prípade, pri preusporiadaní cyklickými posunmi je potrebné vygenerovať parametre posunov – ich veľkosti. PN postupnosti teda majú v obrazových šifrovacích algoritmoch významnú úlohu [3].

Ku jedným z najčastejšie používaných zobrazení patrí logistické zobrazenie (LZ), ktoré spopularizoval v 70. rokoch minulého storočia R. May [4]. Spočiatku sa o tomto pomerne jednoduchom zobrazení uvažovalo ako o vhodnom nástroji na generovanie PN postupností [5], no neskôr sa poukázalo na značné nedostatky tohto riešenia [6]. S narastajúcim množstvom publikácií, ktoré bližšie analyzovali vlastnosti LZ [3] však začali vznikať aj riešenia, ktoré nedostatky LZ vedeli potlačiť do takej miery, že je možné toto riešenie efektívne používať v oblasti obrazových šifrovacích algoritmov [7], [8]. V konečnom dôsledku je možné tvrdiť, že LZ a jeho vlastnosti sú v súčasnosti dostatočne preskúmané na rozdiel od novších viacrozmerných zobrazení [9].

Niektoré algoritmy využívajúce LZ však zohľadňujú iba časť záverov vyplývajúcich z analýzy, napr. obmedzia sa na využitie určitého intervalu parametra zobrazenia namiesto toho, aby skúmali možnosti využitia iných intervalov [10]. Takisto sú niektoré všeobecné závery nevhodne implementované – veľa riešení zohľadňuje iba maximálne hodnoty odhadnutých Ljapunovových exponentov (LE) bez ohľadu na ich minimálne hodnoty, ktoré sú pre použitie v obrazových šifrovacích algoritmoch kritické [11], [12].

Tento príspevok sa v ďalších kapitolách bude venovať popisu LZ, jeho rovnice a z nej vyplývajúcich vlastností zobrazenia. Takisto bude uvedený aj postup odhadu LE vo všeobecnosti a aj v prípade použitia LZ. Po demonštrácii vplyvu odhadnutých LE na vlastnosti postupnosti generovanej LZ bude uvedený aj stručný záver vyplývajúci z nášho výskumu v tejto oblasti.

II. LOGISTICKÉ ZOBRAZENIE

Logistické zobrazenie (LZ) je jednorozmerné chaotické zobrazenie, ktoré využíva parameter $r \in (0; 4)$ a počiatočnú hodnotu $x_0 \in (0; 1)$ na generovanie iterátov $x_n \in (0; 1)$, kde poradové číslo iterátu $n = 1, 2, 3, \dots, N$ a N je celkový počet iterátov. Tieto iteráty je možné ďalej spracovať a napr. vhodnou kvantizáciou z nich získať prvky PN postupnosti [8], [12]. Rovnicu LZ je možné zapísať ako (1):

$$x_{n+1} = r \cdot x_n(1 - x_n). \quad (1)$$

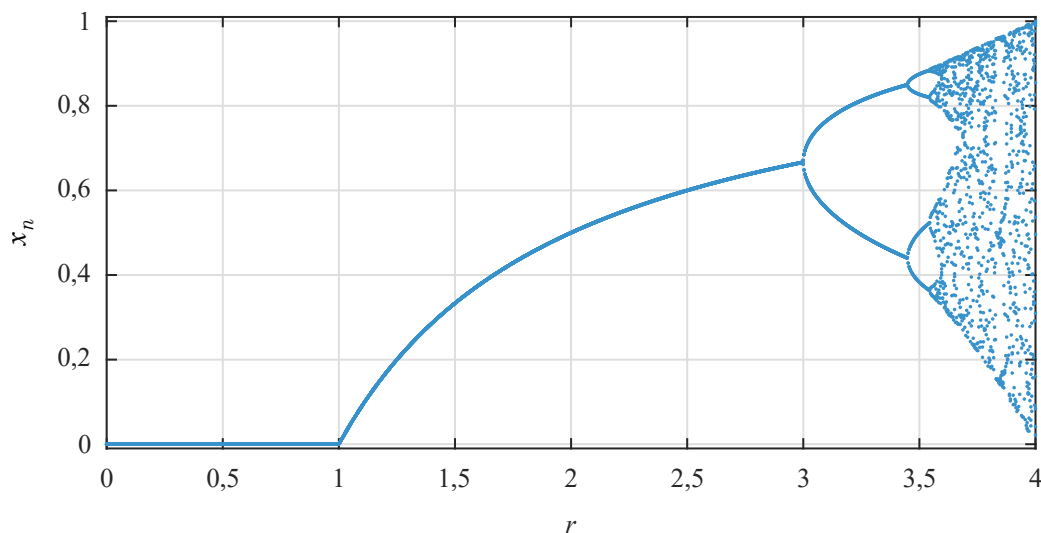
Analýzou LZ sa zistilo, že výpočet iterátov pomocou (1) môže vytvoriť jednoducho odhaliteľné závislosti medzi ich hodnotami na začiatku postupnosti [4]. Preto sa určitý počet iterátov nepoužíva na generovanie PN postupnosti, ale iba kvôli zmene počiatočnej hodnoty. Tieto iteráty sú počítané v tzv. prechodnej dobe a jej dĺžka je zväčša 100 alebo 1000 iterátov [4], [12].

Pri práci s iterátmi LZ treba zdôrazniť, že toto zobrazenie vychádza z diferenciálnej rovnice [4], teda vypočítané iteráty sú desatinné čísla. Pre praktické účely ako je napr. použitie pri šifrovaní obrazov je potrebné hodnoty týchto iterátov diskretizovať – používa sa najmä kvantizácia, pri ktorej však dochádza ku tzv. dynamickej degradácii chaosu [8], [13]. Tento nepriaznivý jav vzniká kvôli konečnému počtu desatinných čísel pri používaných dátových typoch, zatiaľ čo v spojitých oblastiach je počet týchto čísel nekonečný.

Dátové typy pre desatinné čísla sú definované v štandarde IEEE 754 [14]. Vo výpočtových prostrediach ako napr. MATLAB je najpresnejším dátovým typom pre reprezentáciu desatinných čísel typ *double precision*, ktorý vyjadruje desatinnú časť čísla 52 bitmi. Iterát je teda reprezentovaný s presnosťou na $\log_{10} 2^{52} = 15,6536$ desatinných miest. V prípade, ak by sa dvojica iterátov líšila na šestnástom desatinnom mieste, nie je možné zaručiť, že budú v dátovom type *double precision* reprezentované inými hodnotami.

Takéto obmedzenie negatívne vplýva na generované postupnosti iterátov. Chaotické zobrazenia sú známe svojou citlivosťou na vstupné hodnoty a tak aj malá zmena hodnoty iterátu môže významne ovplyvniť dynamické vlastnosti zobrazenia [4], [15]. Kvôli tomu sa môžu v postupnosti vyskytnúť periodické oblasti, kde sa s určitou periódou opakuje rovnaká postupnosť iterátov, alebo tzv. pevné body zobrazenia, pri ktorých sa hodnota nasledujúcich iterátov nemení. Takéto situácie sú pri použití LZ ako generátora PN postupnosti nežiaduce a je potrebné ich potlačenie [12].

Základné vlastnosti LZ je možné jednoducho demonštrovať prostredníctvom bifurkačného diagramu. Ten zobrazuje priebeh postupnosti iterátov v závislosti na hodnote parametra r . Bifurkačný diagram LZ s 1000 vzorkami pre interval $r \in (0; 4)$ a prechodnou dobou s dĺžkou 1000 iterátov pre každú vzorku je ilustrovaný na Obr. 1.

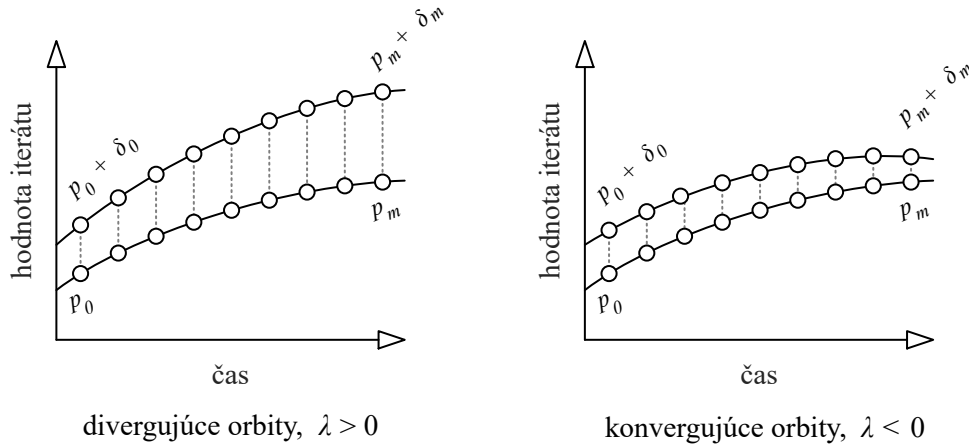


Obr. 1 Bifurkačný diagram logistického zobrazenia

Hlavný záver vyplývajúci z bifurkačného diagramu LZ je skutočnosť, že s rastúcou hodnotou parametra r sa zväčšuje počet rôznych trajektórií hodnôt iterátov. V prípade $r < 3$ je trajektória jedna a hodnoty po sebe nasledujúcich iterátov sú pomerne jednoducho odhaliteľné. V bode $r \sim 3$ nastáva prvá bifurkácia a hodnoty iterátov oscilujú medzi dvoma trajektóriami. Pre praktické aplikácie v oblasti obrazových šifrovacích algoritmov sa využíva $r \geq 3,56995$, pričom tento bod je označovaný ako “počiatok chaosu” [4], [15].

III. LJAPUNOVOVE EXPONENTY A ICH ODHAD

Pre kvantifikáciu vlastností chaotických zobrazení sa dajú používať aj zložitejšie metriky ako napr. Ljapunovove exponenty (LE), označené ako λ . Pointa pri odhade LE spočíva v tom, že sa zvolia dva počiatočné body p_0 a $p_0 + \delta_0$, kde δ_0 je počiatočná odchýlka týchto bodov. Pri zachovaní konštantných parametrov (v prípade LZ hodnota r) sa skúma, či sa odchýlka iterovaním zobrazenia zväčšuje, resp. či sa znižuje. Body po m iteráciách sa označujú p_m a $p_m + \delta_m$, kde δ_m je konečná odchýlka. Priebehy hodnôt iterátov sa nazývajú orbitami, teda zistí sa, či orbity divergujú (vzájomne sa vzdávajú) alebo konvergujú (vzájomne sa približujú). Tieto situácie sú ilustrované na Obr. 2.



Obr. 2 Divergujúce a konvergujúce orbity pri odhade Ljapunovových exponentov

Vo všeobecnosti sa zobrazenie považuje za chaotické v prípade, ak $\lambda > 0$ [15]. Okrem dvoch situácií ilustrovaných na Obr. 2 môže nastať aj prípad, kedy $\lambda = 0$. Pri tejto hodnote orbity nedivergujú, ani nekonvergujú, ale ich odchýlka je konštantná.

Hodnoty LE λ sa dajú odhadnúť pomocou vzťahu (2):

$$\lambda = \frac{1}{m} \cdot \left| \frac{\delta_m}{\delta_0} \right| = \frac{1}{m} \cdot \ln |(f^m)'(x_0)| \sim \lim_{m \rightarrow \infty} \frac{1}{m} \cdot \sum_{i=0}^{m-1} \ln |f'(x_i)|, \quad (2)$$

kde m je počet vzoriek, zátvorky $|a|$ označujú absolútnu hodnotu čísla a , $\ln(b)$ je prirodzený logaritmus čísla b , $(f^m)'$ reprezentuje hodnotu funkcie f v bode m , $(f^m)'(x_0)$ je derivácia f_m vzhľadom na x_0 a i je poradové číslo iterátu.

Z (2) vyplýva, že pri konečnom počte vzoriek ($m < \infty$) sa nejedná o presný výpočet hodnoty λ , ale iba o jej odhad. Presnosť odhadu závisí na voľbe m , v praktických aplikáciách sa zvyčajne používa hodnota $m = 1000$ vzoriek.

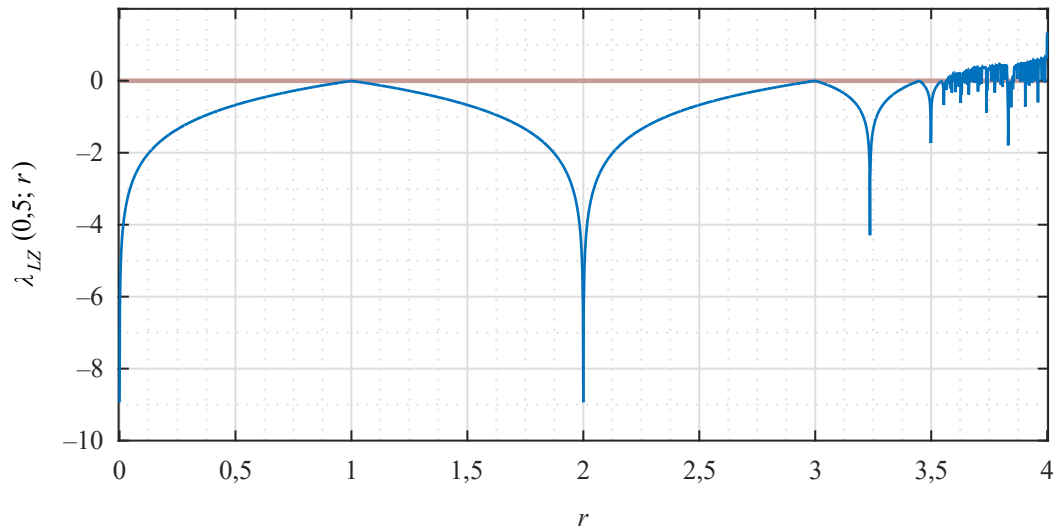
Pre potreby odhadu hodnoty LE LZ λ_{LZ} je možné (2) upraviť dosadením (1) za premennú f . Po úprave získavame (3):

$$\lambda_{LZ} \sim \lim_{m \rightarrow \infty} \frac{1}{m} \cdot \sum_{i=0}^{m-1} \ln |r(1 - 2x_i)|. \quad (3)$$

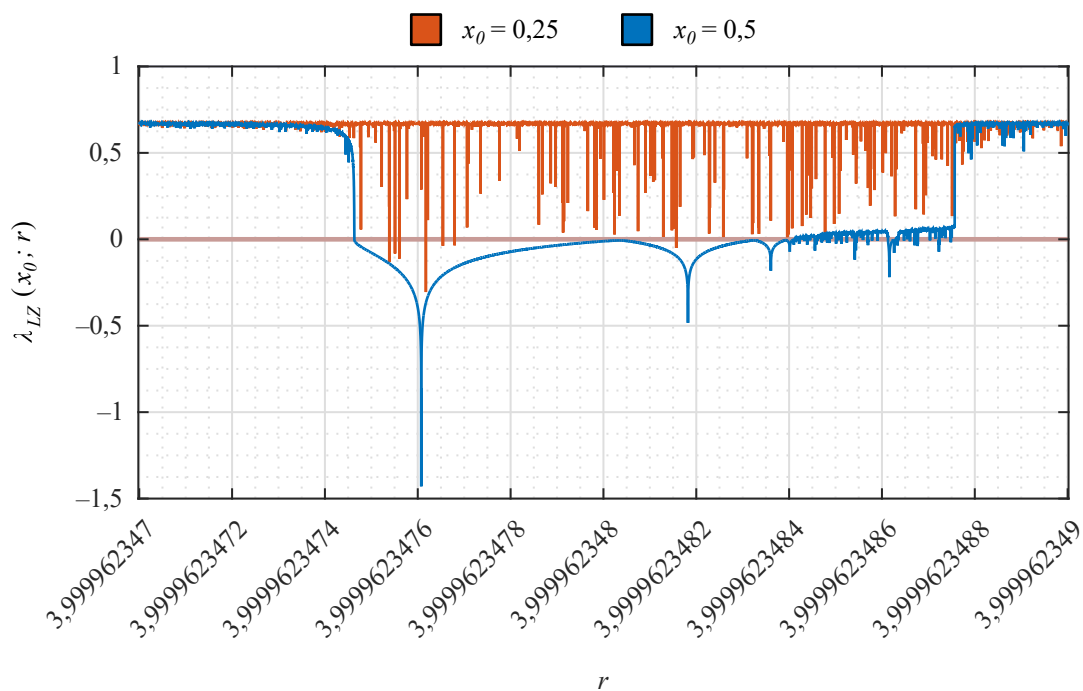
Z (3) je zreteľné, že hodnota λ_{LZ} závisí od hodnoty parametra zobrazenia r . Takisto závisí aj od hodnôt iterátov x_i , ktoré ovplyvňuje počiatočná hodnota x_0 [12], [16]. Kvôli týmto skutočnostiam je potrebné v oblasti obrazových šifrovacích algoritmov presne uvádzať parametre použité pri odhade LE, nie len výsledné hodnoty LE.

Priebeh odhadnutých hodnôt LE pre LZ λ_{LZ} je znázornený na Obr. 3. Na získanie tohto priebehu so 40000 vzorkami na intervale $r \in (0; 4)$ sa použili hodnoty $x_0 = 0,5$ a $m = 1000$ vzoriek. Dĺžka prechodnej doby bola 1000 iterátov. Červená čiara symbolizuje hodnotu $\lambda = 0$, pre hodnoty λ_{LZ} nad ňou je možné zobrazenie považovať za chaotické.

Na Obr. 3 je viditeľné, že LZ je možné podľa hodnôt λ_{LZ} považovať za chaotické od tzv. "počiatku chaosu" ($r \geq 3,56995$). Avšak aj pri vyšších hodnotách parametra r môže byť odhadnutá hodnota λ_{LZ} záporná. Na túto situáciu spolu s vplyvom rozličných počiatočných hodnôt x_0 na odhadnuté hodnoty λ_{LZ} poukazuje Obr. 4. V tomto prípade bola dĺžka prechodnej doby 1000 iterátov, hodnota m bola 1000 vzoriek a vykreslené priebehy mali počet vzoriek ekvivalentný ku 4×10^{13} vzorkám pre interval $r \in (0; 4)$.



Obr. 3 Priebeh odhadnutých hodnôt Ljapunovových exponentov pre logistické zobrazenie



Obr. 4 Detail priebehu odhadnutých hodnôt Ljapunovových exponentov pre logistické zobrazenie

Praktický význam záporných hodnôt λ_{LZ} je možné ukázať aj jednoduchým príkladom. Pri $x_0 = 0,5$ bola v bode $r = 3,99996234761$ hodnota λ_{LZ} odhadnutá na približne $-1,42814$. Vlastnosti vygenerovanej postupnosti iterátov ukazuje Tabuľka 1.

Z Tabuľky 1 je zjavné, že pri použití uvedených hodnôt r a x_0 je generovaná postupnosť periodická s periódou 9 iterátov. Táto situácia nastáva kvôli dynamickej degradácii chaosu – hodnota iterátu x_9 je natoľko podobná počiatkovej hodnote x_0 , že pri použití dátového typu double precision sú obidve tieto hodnoty reprezentované rovnakým číslom. Pre PN postupnosti je výskyt periodicity neprípustný [6], [12].

Niektoré práce sa sústreďujú na maximálne hodnoty LE, ktoré charakterizujú vlastnosti obrazových šifrovacích algoritmov v najlepšom prípade. Avšak stojí za zváženie, či nie je dôležitejšie uvádzať minimálne hodnoty LE, ktoré vyjadrujú vlastnosti obrazových šifrovacích algoritmov v najhoršom prípade. Minimálna hodnota LE sa dá do istej miery vylepšiť vhodným výberom intervalu parametra r . Prílišné zmenšenie intervalu však negatívne ovplyvňuje počet kľúčov obrazových šifrovacích algoritmov [12].

Tabuľka 1
 Periodické vlastnosti generovanej postupnosti iterátov LZ pri zápornej hodnote λ_{LZ}

$x_0 = 0,5; r = 3,99996234761; \lambda_{LZ}(x_0; r) \sim -1,42814$	
$x_1 = 0,999990586901875$	$x_{10} = x_1 = 0,999990586901875$
$x_2 = 0,000037651683653$	$x_{11} = x_2 = 0,000037651683653$
$x_3 = 0,000150599646393$	$x_{12} = x_3 = 0,000150599646393$
$x_4 = 0,000602302194975$	$x_{13} = x_4 = 0,000602302194975$
$x_5 = 0,002407735043706$	$x_{14} = x_5 = 0,002407735043706$
$x_6 = 0,009607660983953$	$x_{15} = x_6 = 0,009607660983953$
$x_7 = 0,038061057061643$	$x_{16} = x_7 = 0,038061057061643$
$x_8 = 0,146448273443029$	$x_{17} = x_8 = 0,146448273443029$
$x_9 = 0,500000000000973$	$x_{18} = x_9 = 0,500000000000973$
$x_9 \sim x_0$...

IV. ZÁVER

Tento príspevok sa zaoberal logistickým zobrazením a kvantifikáciou jeho chaotických vlastností pomocou odhadu Ljapunovových exponentov. Boli vysvetlené rôzne aspekty ich odhadu, vďaka ktorým môže prísť ku nezrovnalostiam pri porovnávaní obrazových šifrovacích algoritmov. Z nášho výskumu vyplýva jeden stručný záver – je potrebné používať intervaly parametra r , pre ktoré je odhadnutá hodnota Ljapunovových exponentov kladná.

V budúcnosti sa náš výskumný tím chce venovať zefektívneniu odhadu Ljapunovových exponentov do takej miery, aby sa experimentálne dali odhadnúť ich hodnoty s presnosťou dátového typu double precision. V súčasnosti je možné vo výpočtovom prostredí MATLAB dosiahnuť rozlíšenie 4×10^{13} vzoriek, pri vyššom rozlíšení dochádza k vyčerpaniu alokovanej pamäte.

POĎAKOVANIE

Táto práca bola podporená výskumným projektom FEI-2022-84 “Techniky spracovania dát vo vysokorýchlostných prenosových systémoch”.

LITERATÚRA

- [1] J. Fridrich, Symmetric Ciphers Based on Two-Dimensional Chaotic Maps, *Intl. J. of Bifurcation and Chaos*, 1998, roč. 8, č. 6, s. 1259–1284. ISSN: 0218-1274. DOI: 10.1142/S021812749800098X.
- [2] FIPS 197: Advanced encryption standard. DOI: 10.6028/NIST.FIPS.197.
- [3] M. Preishuber, T. Hütter, S. Katzenbeisser a kol., Depreciating Motivation and Empirical Security Analysis of Chaos-based Image and Video Encryption, *IEEE Transactions on Information Forensics and Security*, 2018, roč. 13, č. 9, s. 2137-2150. ISSN: 1556-6013. DOI: 10.1109/TIFS.2018.2812080.
- [4] R. May, Simple Mathematical Models with Very Complicated Dynamics, *Nature*, 1976, roč. 261, č. 5560, s. 459–467. ISSN: 0028-0836. DOI: 10.1038/261459a0.
- [5] S. C. Phatak, S. S. Rao, Logistic Map: A Possible Random-Number Generator, *Physical Review E*, 1995, roč. 51, č. 4, s. 3670–3678. ISSN: 2470-0045. DOI: 10.1103/PhysRevE.51.3670.
- [6] D. Arroyo, G. Alvarez, V. Fernandez, On the Inadequacy of the Logistic Map for Cryptographic Applications, v *Proc. of 10th Spanish Meeting on Cryptology and Information Security*. Salamanca (Španielsko), 2008, s. 1–6.
- [7] L. Liu, S. Miao, A New Image Encryption Algorithm Based on Logistic Chaotic Map with Varying Parameter, *SpringerPlus*, 2016, roč. 5, č. 1, s. 289-300. ISSN: 2193-1801. DOI: 10.1186/s40064-016-1959-1.
- [8] J. Oravec, J. Turán, L. Ovseník a kol., Chaotic Image Encryption Algorithm Robust against the Phase Space Reconstruction Attacks, *Acta Polytechnica Hungarica*, 2019, roč. 16, č. 3, s. 37-57. ISSN: 1785-8860. DOI: 10.12700/aph.16.3.2019.3.3.
- [9] Z. Li, A Novel Plaintext-related Image Encryption Scheme Using Hyper-chaotic System, *Nonlinear Dynamics*, 2018, roč. 94, č. 4, s. 1319-1333. ISSN: 0924-090X. DOI: 10.1007/s11071-018-4426-4.
- [10] W. K. Lee, R. C. Phan, W.-S. Yap a kol., SPRING: A Novel Parallel Chaos-based Image Encryption Scheme, *Nonlinear Dynamics*, 2018, roč. 92, č. 3, s. 575-593. ISSN: 0924-090X. DOI: 10.1007/s11071-018-4076-6.
- [11] B. Ramalingam, D. Ravichandran, A. A. Annadurai a kol., Chaos Triggered Image Encryption – a Reconfigurable Security Solution, *Multimedia Tools and Applications*, 2017, roč. 77, č. 10, s. 11669-11692. ISSN: 1380-7501. DOI: 10.1007/s11042-017-4811-x.
- [12] J. Oravec, L. Ovseník, J. Turán a kol., Mitigating Drawbacks of Logistic Map for Image Encryption Algorithms, *Computing and Informatics*, 2020, roč. 39, č. 6, s. 1249-1280. ISSN: 1335-9150. DOI: 10.31577/cai2020_6_1250.
- [13] S. Li, G. Chen, X. Mou, On the Dynamical Degradation of Digital Piecewise Linear Chaotic Maps, *International Journal of Bifurcation and Chaos*, 2005, roč. 15, č. 10, s. 3119-3151. ISSN: 0218-1274. DOI: 10.1142/S0218127405014052.
- [14] IEEE 754-2019: IEEE Standard for Floating-point Arithmetic. DOI: 10.1109/IEEESTD.2019.8766229.
- [15] J. Gleick, *Chaos: Making a New Science*. Londýn: Vintage Books, 1998, 380 s. ISBN: 978-07-4938-606-1.
- [16] E. Ibarra Olivares, R. Vázquez-Medina, M. Cruz-Irison a kol., Numerical Calculation of the Lyapunov Exponent for the Logistic Map, v *Proc. of 12th International Conference on Mathematical Methods in Electromagnetic Theory*. Odessa (Ukraine), 2008, s. 409–411. DOI: 10.1109/MMET.2008.4581011.

Sebestačnosť elektroenergetiky Slovenska

Erik ANDRAŠKO, Dušan MEDVEĎ

Katedra elektroenergetiky, Fakulta elektrotechniky a informatiky,
Technická univerzita v Košiciach, Slovenská republika

erik.andrasko@student.tuke.sk, dusan.medved@tuke.sk

Abstrakt — Tento príspevok sa zaoberá problematikou sebestačnosti na Slovensku, pričom konkrétne rieši možnosti výroby elektrickej energie pomocou veternej energie na území Slovenska. V úvode je predstavený aktuálny stav danej problematiky a celkové možnosti využitia veternej energie na Slovensku, vzhľadom na súčasné podmienky. Popísaný je aj projekt prvej veľkej veternej elektrárne, ktorý môže zvýšiť elektroenergetickú sebestačnosť a bezpečnosť, no zároveň je spomenutý aj finančný aspekt a možné komplikácie pri procesoch schvaľovania takýchto projektov.

Kľúčové slová — elektroenergetika, sebestačnosť, veterná energia, vietor

I. ÚVOD

Byť v dnešnej neistej dobe sebestačný v akoľvek odvetví je nenahraditeľná výhoda. Výnimkou nie je ani elektroenergetika a jej sebestačnosť na úrovni štátu, regiónov, miest či obcí. Neustály rast cien energií a najmä zhoršujúci sa stav klimatej otázky ovplyvňuje aj spôsoby výroby elektrickej energie, kedy sa stále viac a viac do popredia dostáva vďaka vyspelým ekonomikám – najmä na európskom kontinente takzvaná „zelená energia“, a teda elektrická energia vyrobená z obnoviteľných zdrojov energie (OZE). Vo všeobecnosti sa dá povedať, že cesta sebestačnosti elektroenergetiky na Slovensku vedie práve cez obnoviteľné zdroje energie vzhľadom na to, že Slovenská republika dlhodobo patrí medzi krajiny s vysokou dovozovou závislosťou a väčšinu primárnych energetických zdrojov (PEZ) dováža, a tým pádom nespĺňa jednu zo základných podmienok, ktorá je definovaná pojmom sebestačnosť. V tomto príspevku boli bližšie preskúmané možnosti k dosiahnutiu elektroenergetickej sebestačnosti na Slovensku vďaka veternej energii.

II. ENERGIA VETRA

Veterná energia má pôvod v energii Slnka. Zemský povrch ohrievajú slnečné lúče s rôznou intenzitou, v dôsledku čoho dochádza k teplotným a tlakovým rozdielom. Nerovnomerným zohrievaním vzdušných mäs pri zemskom povrchu sa vytvárajú tlakové výše a tlakové níz. Vietor vzniká vplyvom tlakových rozdielov a vane v zásade od tlakovej výše k tlakovej níz. Miestna rýchlosť vetra závisí od tvaru zemského povrchu, so vzdialenosťou od mora klesá, s nadmorskou výškou rastie; mení sa s časom a tiež v závislosti od meteorologickej situácie [1].

Najväčší vplyv na využívanie veternej energie má rýchlosť vetra. Energiu vetra je možné definovať ako kinetickú energiu pohybujúcej sa častice vzťahom [2]:

$$E_k = \frac{1}{2} \cdot m \cdot v^2 \quad [\text{J}] \quad (1)$$

kde: m hmotnosť častice (t.j. vzduchu; pohybujúceho sa média)
 v rýchlosť vetra

Pre hmotnosť častice ďalej platí:

$$m = \rho \cdot V = \rho \cdot A \cdot s \quad [\text{kg}] \quad (2)$$

kde: ρ hustota telesa alebo častice
 A plocha, ktorou vzduch prechádza
 s dráha častice

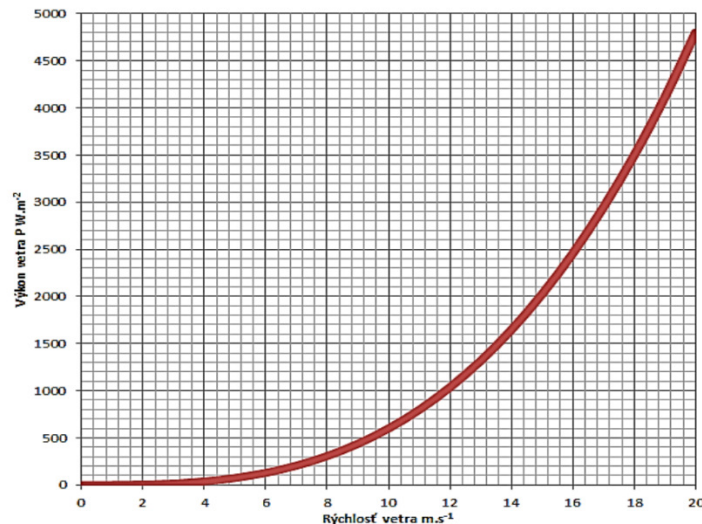
Z definície kinetickej energie pohybujúceho sa telesa môžeme určiť vzorec pre výpočet výkonu vetra P_w , ktorý pôsobí na 1 m^2 plochy [2]:

$$P_w = \frac{E_k}{A \cdot t} = \frac{1}{2} \cdot \rho \cdot \frac{A \cdot s}{A \cdot t} \cdot v^2 = \frac{1}{2} \cdot \rho \cdot v^3 \quad [\text{W}] \quad (3)$$

Pričom ďalej platí:

$$v = \frac{s}{t} \quad [\text{m} \cdot \text{s}^{-1}] \quad (4)$$

Podľa vyššie zadefinovanej rovnice (3) sa dá povedať, že výkon vetra je závislý od hustoty vetra ρ a kvadrátu rýchlosti vetra v , čo reprezentuje aj obrázok 1, z ktorého vyplýva, že rýchlosť vetra podstatne ovplyvňuje jeho výkon [2].



Obr. 1 Závislosť rýchlosti vetra v od výkonu vetra P_w [2]

III. VÝROBA ELEKTRICKEJ ENERGIE VYUŽITÍM ENERGIE VETRA

Veterné elektrárne premieňajú energiu prúdenia vzduchu na elektrinu. Vietor sa oprie o vhodné nastavené krídla rotora turbíny a roztočí ich. Točivá sila rotora sa prenáša cez prevodovku alebo priamo na elektrický generátor, ktorý produkuje elektrinu. Generátor môže byť asynchrónny, synchronný alebo jednosmerný. Generátory s bežným inštalovaným výkonom 500 kW až 3 MW môžu pracovať s konštantnými alebo aj premenlivými otáčkami [1].

Veterné elektrárne vybavené asynchrónnymi generátormi sú určené pre paralelný chod so sieťou. Je vhodné doplniť ich mikroprocesorovým riadiacim systémom, ktorý umožňuje diaľkové ovládanie a sledovanie chodu. Jednotky so synchronnými generátormi môžu pracovať nielen paralelne so sieťou, ale i nezávisle od nej. Môžu sa využívať napríklad ako zdroje, ktoré zabezpečujú potrebnú dodávku elektrickej energie do elektrickej siete pri štarte z tmy. Prúd jednosmerného generátora sa môže používať napríklad na elektrolýzu vody a nosičom energie je potom vodík. Z dôvodu nízkych a premenlivých otáčok veternej turbíny je výhodnejšie používať asynchrónne generátory a meniče s jednosmerným medziobvodom, ktorými sa získané napätie najprv usmerní a potom invertuje na požadované striedavé napätie o frekvencii 50 Hz. Na jednej strane, tento statický usmerňovač s invertorom zvyšuje kapitálové náklady, na druhej strane umožňuje vyrábať veľmi kvalitné napätie a zjednodušuje paralelnú prevádzku so sieťou [1].

Veterné turbíny s vertikálnou osou rotácie sú výhodné pri vysokých a stredných rýchlostiach vetra. Pri prekročení optimálnych rýchlostí vetra rýchlo klesá ich účinnosť a dochádza k poklesu výkonu. Sú nezávislé od smeru vetra a lopatky majú nízko pri zemi. Veterné turbíny s vertikálnou osou majú vyššie hranice minimálnych štartovacích rýchlostí (nutnosť pomocného štartéra) a nižšiu účinnosť. Vyžadujú aj väčšiu zastavanú plochu. Veterné turbíny s horizontálnou osou rotácie sú v súčasnej dobe najrozšírenejšie. Najvyššie využitie výkonu je možné dosiahnuť dvoj- a trojlístovými vrtuľami. Aby mohla vrtuľa čo najlepšie zachytiť energiu vetra, je hlava veternej elektrárne, tzv. gondola, umiestnená na stožiarí otáčavo. Na nasmerovanie osí vrtule v smere vetra slúži postranné veterné koleso alebo elektromotor [1].

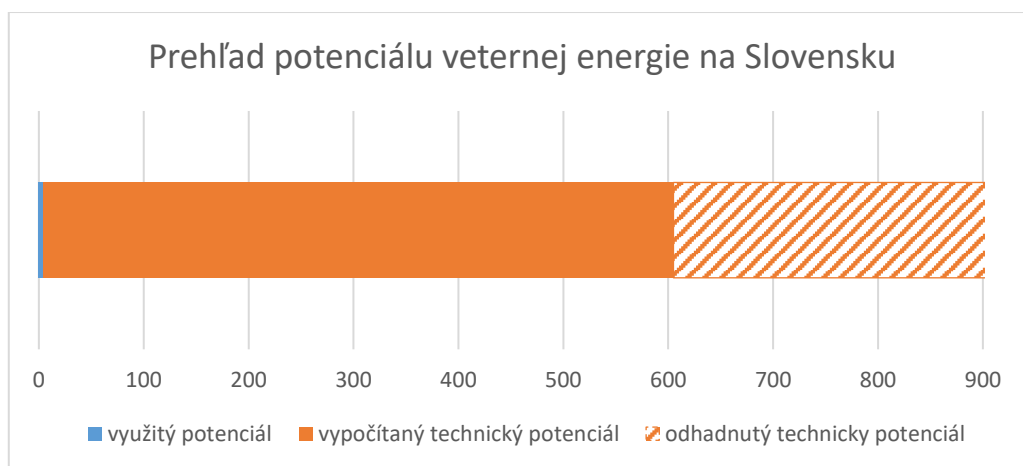
Výkon veternej turbíny je daný vzťahom [3]:

$$P = \frac{1}{2} \cdot \rho \cdot c_p \cdot S \cdot v^3 \quad [\text{W}] \quad (5)$$

- kde: ρ hustota vzduchu
 c_p je súčiniteľ výkonu, ktorý je daný pomerom skutočného a teoretického výkonu vetra. Horná hranica súčiniteľa výkonu je tzv. Betzov limit, ktorý určuje maximálny výkon veternej turbíny v pomere 16/27. Je teda zrejme, že maximálny možný teoretický výkon veternej turbíny je 59,3 % z teoretického výkonu vetra.
 S je plocha, ktorú opisujú lopatky turbíny (plocha rotora)
 v je rýchlosť prúdenia vzduchu

IV. ANALÝZA STAVU PROBLEMATIKY VETERNEJ ENERGIE NA SLOVENSKU

Naposledy bol technický využiteľný potenciál výroby elektrickej energie z veternej energie vyčíslený v roku 2003 a činil 605 GWh/rok. Odvtedy tento potenciál nebol v dostatočnej miere ďalej zmapovaný, neprebehla žiadna jeho aktualizácia a chýba teda aktuálne číslo technického potenciálu vzhľadom aj k novým technológiám v oblasti veternej energie, ktoré za 19 rokov pribudli. Taktiež je potrebné poznamenať, že potenciál bol vypočítaný za predpokladu použitia turbín s inštalovaným výkonom 0,5 MW až 1 MW. Dnes už existujú turbíny s rádovo vyššími inštalovanými výkonmi, a preto je potrebné uvažovať aj s vyšším technickým potenciálom, pričom nie je vylúčené zdvojnásobenie technického využiteľného potenciálu z roku 2003. Keďže by to bolo len hádanie, berme do úvahy, že vďaka spomenutým aspektom je dnes technický potenciál o polovicu vyšší ako v roku 2003 a teda je na úrovni cca 900 GWh/rok [3].



Obr. 2 Graf zobrazujúci využitý a vypočítaný, respektíve odhadnutý technický potenciál veternej energie na území Slovenska

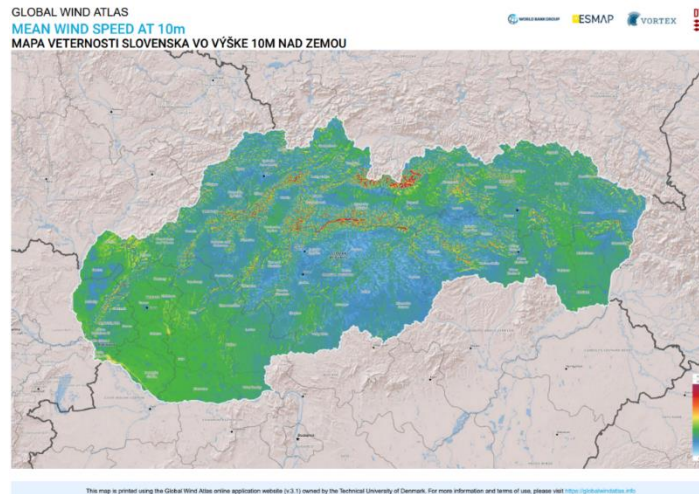
Tu je veľmi jasne vidieť, ako sa veterná energia na Slovensku vôbec nevyužíva, keďže pri vypočítanom technickom potenciáli 605 GWh/rok sa v roku 2020 podľa OKTE vyrobilo len 3,925 GWh, čo je v prepočte na percentá len 0,649%. V prípade odhadnutého potenciálu je to len 0,436 %, čo v spojení s pojmom elektroenergetická sebestačnosť dáva veľký priestor na realizáciu projektov k jej dosiahnutiu.

Pri pohľade na mapu veternosti vo výške 10 m nad zemou je možné pozorovať, že potenciál sa nachádza na Podunajskej nížine a na Záhorí (v menšine aj Zemplín), keďže len tieto oblasti dosahujú hodnoty technického minima, čo je viac ako 3 m/s, no stále je to nedostatočná rýchlosť vetra pre zmysluplnú výrobu elektrickej energie. Z ekonomického hľadiska je potrebné mať vo výške 10 m rýchlosť vetra aspoň 5 m/s, aby prevádzka veternej elektrárne mala zmysel. Červené miesta na mape znázorňujú vysoké hrebene pohorí, kde je však neprípustná akákoľvek výstavba, nie to ešte veterné turbíny. Ako najmenej perspektívna oblasť pre využitie veternej energie sa javí juh stredného Slovenska, Horehronie a Kysuce, kde priemerná rýchlosť vetra v desiatich metroch je len na úrovni 1,5 až 2,5 m/s.

Je dobre známe, že so zvyšujúcou sa výškou rastie aj rýchlosť vetra. Preto aj na prvý pohľad nezaujímavých krajinách, z pohľadu výroby elektrickej energie z vetra (akou Slovensko bezpochyby je), sa inštalovaním veľmi vysokých veterných turbín, dajú dosiahnuť zaujímavé hodnoty produkcie elektrickej energie. Pre úplnosť informácií, rýchlosť vetra vo výške desiatich metrov sa považuje za referenčnú rýchlosť, podľa ktorej sa orientačne počíta rýchlosť vetra v_p v požadovanej výške h_p podľa vzorca [3]:

$$v_p = v_{ref} \cdot \left(\frac{h_p}{h_{ref}} \right)^\alpha \quad [\text{m} \cdot \text{s}^{-1}] \quad (6)$$

- kde v_{ref} je rýchlosť vetra vo výške 10 m [$\text{m}\cdot\text{s}^{-1}$]
 h_p je výška bodu, v ktorom sa meria rýchlosť vetra nad terénom [m]
 h_{ref} je výška bodu nad terénom [10 m]
 α je koeficient drsnosti terénu (0,1 až 0,45, podľa terénu)



Obr. 3 Mapa Slovenska s vyznačenou veternosťou vo výške 10 m od povrchu (zdroj: globalwindatlas)

Na veternej mape Slovenska boli zvolené 3 lokality z ktorých sa vybrali údaje z veternostných máp vo výške 10 m a 100 m. Následne sa pomocou výpočtu znázorneného vyššie vypočítala rýchlosť vo výške 100 m, ktorá sa porovnávala s rýchlosťou vetra podľa mapy, uvedenej na globalwindatlas:

Pri meste Šaľa na Podunajskej nížine 3,59 m/s (10 m) a 6,36 m/s (100 m)

$$v_p = 3,59 \cdot \left(\frac{100}{10}\right)^{0,25} = 6,38 \text{ m}\cdot\text{s}^{-1} \quad (7)$$

Pri meste Skalica na Záhorí 4,97 m/s (10 m) a 7,59 m/s (100 m)

$$v_p = 4,97 \cdot \left(\frac{100}{10}\right)^{0,25} = 8,84 \text{ m}\cdot\text{s}^{-1} \quad (8)$$

Pri meste Poltár na strede južného Slovenska 1,67 m/s (10 m) a 3,21 m/s (100 m)

$$v_p = 1,67 \cdot \left(\frac{100}{10}\right)^{0,25} = 2,97 \text{ m}\cdot\text{s}^{-1} \quad (9)$$

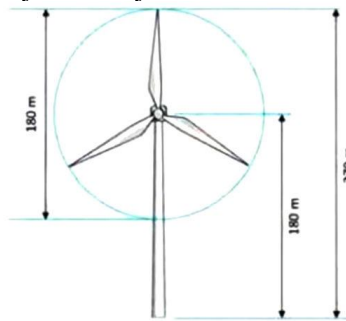
Z týchto výpočtov je možné pozorovať, že pri prípadnej realizácii projektov veterných elektrární je dôležité vybrať si vhodnú lokalitu na území Slovenska. Podľa medzinárodnej stupnice veternostných tried IEC (Iowa Energy Center), patrí Slovensko do veternostnej triedy 2 až 3. To znamená, že pre efektívnu výrobu elektrickej energie z vetra je potrebný väčší priemer rotora turbíny, ktorá musí byť zároveň inštalovaná na vyšších stožiaroch [4].

V. MOŽNOSTI PRE DOSIAHNUTIE SEBETAČNOSTI VYUŽITÍM VETERNEJ ENERGIE

Momentálne je analyzovaný zámer prvej veľkej veternej elektrárne v okrese Skalica, ktorá by mala mať inštalovaný výkon 30 MW, respektíve 42 MW, v závislosti či bude realizovaný variant s piatimi veternými turbínami, alebo variant až so siedmimi turbínami. Výkon jednej je dimenzovaný na 6 MW a jej parametre, čo sa týka veľkosti, je možné vidieť na obrázku 4. Ako už bolo spomenuté, aby sa využil potenciál veternej energie na Slovensku, je potrebné použiť turbíny na vysokých stožiaroch, v tomto prípade sa uvažuje s výškou až 180 m a priemer rotora turbíny má mať takisto 180 m (dĺžka jednej lopatky 90 m). To znamená, že celková výška by mala byť až 270 m. V závislosti od realizácie variantov, investícia do takejto veternej elektrárne by mala byť 36 alebo 50,4 miliónov eur, čo pri prepočte na 1 MW vychádza na 1,2 milióna eur. S prevádzkou sa ráta na obdobie 25 rokov[4].

Je potrebné povedať, že v posledných rokoch nevznikla žiadna veľká veterná elektrárňa na Slovensku. Značným spôsobom zaostávame za technickým potenciálom. Je viacero príčin, kvôli ktorým nie je spôsob výroby elektrickej energie z vetra preferovaný. Najväčším problémom je nesmierne komplikovaný byrokratický proces schvaľovania a povoľovania realizácie veternej elektrárne. Taktiež neaktuálne dáta o veternosti na Slovensku neprispievajú k ľahšej a rýchlejšej

realizácii potencionálnej veternej elektrárne. Ďalším problémom môže byť tiež nie úplne pozitívny prístup odbornej verejnosti k veternej energii, čo môže v konečnom dôsledku ovplyvniť aj prístup verejnosti k stavbe takejto veternej elektrárne v ich bezprostrednej blízkosti [3].



Obr. 4 Navrhované rozmery veternej turbíny pri Skalici (zdroj: WSB Invest)

Pri realizácii projektu veternej elektrárne je tiež potrebné zhodnotiť investície na výstavbu a zároveň aj prevádzku počas jej celej životnosti, ktorá je zvyčajne 20 až 25 rokov. Definícia opodstatnenej investície hovorí o tom, že je nutné dosiahnuť návratnosť do konca životnosti veternej elektrárne. Pod nerentabilnou investíciou sa rozumie investícia, ktorá nedosiahne návratnosť v rozmedzí 4 až 6 rokov od samotného spustenia. S uvažovaním nákladov na výrobu elektrickej energie podľa indexu LCOE, ktoré sú podľa podmienok veternej energie na Slovensku na úrovni 0,0823 €/kWh, dá sa vypočítať miera návratnosti myšleného projektu s parametrami 1500 €/kW, 3200 h prevádzky na nominálnom výkone. Takáto elektrárňa by za rok vyprodukovala 3200 kWh elektrickej energie za sumu 263,36 €. Pri uvažovanej životnosti 20 až 25 rokov by za svoju životnosť vyrobila elektrickú energiu za 5267,2 až 6584 €. Aby bola takáto elektrárňa považovaná za návratnú investíciu, výkupná cena elektriny na Slovensku by musela byť na úrovni 0,1604 €/kWh. Realita je taká, že plánovaná maximálna výkupná cena elektrickej energie pre novo inštalované veterne elektrárne je podľa Ministerstva hospodárstva Slovenskej republiky len 0,08498 €/kWh [3].

VI. ZÁVER

Spolu so všetkými vyššie uvedenými skutočnosťami treba brať do úvahy aj fakt, že vzhľadom na nastavenú výkupnú cenu elektrickej energie je pre väčšinu projektov veterných elektrární investícia nerentabilná, a počas životnosti sa nevráti. V momentálnej situácii dáva zmysel jedine spotreba vyrobenej elektrickej energie v mieste jej výroby alebo prípad, kedy nie je hlavnou prioritou návratnosť, ale skôr sa na projekt hľadí ako na zdroj bezemisnej elektrickej energie, ktorý môže dopomôcť k dosiahnutiu sebestačnosti, respektíve aspoň k jej priblíženiu. Záujem investorov o tento druh výroby elektrickej energie v poslednom období rastie. Ak by sa upravila cenová politika výkupu elektrickej energie a zjednodušila by sa byrokratická časť s rôznymi povoleniami, dá sa predpokladať, že záujem o veternú energiu by ešte viac vzrástol. To by znamenalo masívnejšie využívanie ďalšieho obnoviteľného zdroja na výrobu elektrickej energie na území Slovenska.

POĎAKOVANIE

Tento príspevok vznikol vďaka Agentúre na podporu výskumu a vývoja na základe zmluvy č. APVV-19-0576, ďalej Vedeckej grantovej agentúry Ministerstva školstva, vedy, výskumu a športu SR a SAV podporenej grantom VEGA 1/0757/21.

ZDROJE

- [1] Elektrárne. [online]. [cit. 2022-05-15]. Dostupné na internete: <http://www.oze.stuba.sk/wp-content/themes/ObnovitelneZdrojeEnergie/elearning/EENERGETIKA/ELEN-3_2.htm#3.2.3>.
- [2] Kolcun, M., a kol., „Elektrárne“, 1. vyd. Košice: Technická univerzita v Košiciach, 2017. 202 s. ISBN 978-80-553-3119-5.
- [3] „Kvantifikácia reálne využiteľného potenciálu veternej energie na Slovensku“, [online]. [cit. 2022-05-08]. Dostupné na internete: <https://cepa.priateliazeme.sk/images/publikacie/EVS_vystupy/M08_navrh.pdf>.
- [4] „Prvá veľká veterná elektrárňa má vyrásť na západnom Slovensku“. In energie-portal.sk [online]. [cit. 2022-05-16]. Dostupné na internete: <<https://www.energie-portal.sk/Dokument/prva-velka-veterna-elektren-na-zapadnom-slovensku-106722.aspx>>.
- [5] Andraško, E., „Sebestačnosť elektroenergetiky Slovenska“. Bakalárska práca. Technická univerzita v Košiciach, Fakulta elektrotechniky a informatiky, 2022.

Využitie Householderovej transformácie pri spracovaní hraničných blokov objektov

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 možnosťou vylepšenia spracovania obrazu pri použití Householderovej transformácie. Prezentuje sa tu koncepcia spracovania obrazov pomocou diskkrétnej kosínusovej transformácie spolu s Householderovou transformáciou. Ďalej sa v krátkosti venuje princípu činnosti Householderovej algoritmu ortogonalizácie. Pre účely jeho aplikovania na ľubovoľný obraz bola vytvorená aplikácia v prostredí MATLAB. Pomocou nej boli otestované viaceré obrazy, na ktoré boli aplikované okrem spomínanej metódy použitia oboch transformácií naraz, aj iné metódy spracovania obrazu za účelom porovnania. Dosiahnuté výsledky SNR sú zobrazené číselne a taktiež sú porovnané jednotlivé spôsoby transformácie obrazov a ich následnej rekonštrukcie.

Keľúčovú slová — algoritmus ortogonalizácie, Householderova transformácia, hraničný blok objektu

I. ÚVOD

V dnešnej dobe sa na každom kroku stretávame s rôznymi formami médií. Médiá, ako obrazy alebo videá, zohrávajú podstatnú rolu, a tak ich spracovanie je veľmi dôležité. Ľudia sa neustále snažia spracovať tieto médiá tak, aby mali čo najlepšiu kvalitu pri čo najmenšej veľkosti. Práve preto neustále vznikajú nové formáty, ktoré pracujú na iných princípoch a ich cieľom je dosiahnuť práve tieto ciele. Medzi najpoužívanjšie formáty by sme napríklad mohli zaradiť formáty JPEG a MPEG. Oba tieto formáty využívajú pri spracovaní, resp. kompresii diskkrétnu kosínusovú transformáciu (DCT). Príchodom štandardu videokompresie MPEG-4 súčasne vznikla nová metóda kompresie obrazu pomocou výberu vizuálnych objektov z obrazovej scény [1]. Algoritmy, ktoré vo všeobecnosti kódujú jednotlivé videoobjekty namiesto celých obrazových rámcov, podporujú rozvoj aplikácií s obsahovou interakciou. Z vizuálnej scény sa vyberajú vizuálne objekty, ktoré sa môžu kódovať rôznymi technikami. V koncepcii videoobjektových rovín je každá snímka vstupnej videosekvencie segmentovaná do určitého počtu ohraničených oblastí (videoobjektových rovín), pričom táto oblasť môže pokrývať určitú časť snímky, resp. určitý fyzický objekt v scéne. Tento systém je bližší ľudskému vnímaniu obrazu a umožňuje efektívnu reprezentáciu vizuálnych objektov, manipuláciu s nimi, a editovanie v ich dátových tokoch ako aj možnosť škálovania dávajú nové možnosti pre spracovanie obrazu. DCT sa vďaka jej vlastnostiam stala veľmi obľúbenou, ale aj táto transformácia má určité nedostatky, ktoré by sa dali vylepšiť. Medzi takéto nedostatky by sme napríklad mohli zaradiť spracovanie hraničných blokov objektov v obraze. Kvalitu spracovania hraničných blokov by bolo možné vylepšiť použitím Householderovej transformácie. Tento článok je zameraný práve na túto transformáciu a jej kombináciu s DCT pri spracovaní obrazu a jej porovnanie s podobnými metódami spracovania obrazu.

II. PRINCÍP HOUSEHOLDEROVEJ TRANSFORMÁCIE

Householderov algoritmus resp. Householderova transformácia patrí k najvýkonnejším nástrojom pri maticových výpočtoch, umožňuje lepšiu reprodukciu obrazových segmentov [2]. Táto transformácia zohráva významnú úlohu v numerickej lineárnej algebre, jej zaujímavé vlastnosti sú široko používané pri vytváraní efektívnych a stabilných algoritmov. Householderov algoritmus vypočítava všetky ortogonálne bázové funkcie naraz a jeho nevýhodou je najmä časová náročnosť výpočtu, ktorá so zväčšujúcimi sa blokmi v obrazovom segmente rastie [3]. Medzi ďalšie algoritmy ortogonalizácie by sme mohli zaradiť napríklad Gramov-Schmidtov algoritmus, ktorý je citlivejší na numerické chyby, oproti Householderovmu algoritmu je menej

presný a v niektorých prípadoch môže dochádzať k strate ortogonalnosti [4]. Vysokú presnosť Householderov algoritmus dosahuje najmä kvôli práci s ortogonálnymi maticami [5].

Pri Householderovej transformácii sa snažíme získať hornú trojuholníkovú maticu \mathbf{R} , ktorá pozostáva z ortogonálnych koeficientov s rozmerom $n \times n$. Cieľom je tiež vypočítať Householderovu maticu \mathbf{H} rovnakého rozmeru. Riadky tejto matice \mathbf{H} tvoria ortogonálne bázové funkcie. Maticu \mathbf{R} je možné dostať súčinom Householderových matic a vstupnej matice \mathbf{X} . Rozmery matice \mathbf{X} sú taktiež $n \times n$. Maticu \mathbf{R} dostaneme nasledovne [6]

$$\mathbf{H}_{n-1}\mathbf{H}_{n-2} \dots \mathbf{H}_1\mathbf{X} = \mathbf{R} \quad (1)$$

Pre vypočítanie matice \mathbf{R} je najskôr nutné vypočítať Householderove matice $\mathbf{H}_1, \mathbf{H}_2$ až \mathbf{H}_{n-1} .

Maticu \mathbf{H}_1 vypočítame pomocou vzťahu

$$\mathbf{H}_1 = \mathbf{I} - 2\mathbf{w}\mathbf{w}^T \quad (2)$$

kde \mathbf{I} predstavuje jednotkovú maticu a \mathbf{w} normovaný stĺpcový vektor. Tento normovaný stĺpcový vektor \mathbf{w} , vyhovujúci rovnici (2) je daný vzťahom

$$\mathbf{w} = \frac{\mathbf{x}_1 - k\mathbf{e}_1}{\|\mathbf{x}_1 - k\mathbf{e}_1\|} \quad (3)$$

kde \mathbf{e}_1 je jednotkový normovaný vektor.

Maticy \mathbf{H}_1 až \mathbf{H}_{n-1} budú symetrické a ortogonálne matice [7], čo je možné dokázať pomocou nasledujúcich rovníc

$$\mathbf{H}_1^T = \mathbf{I}^T - (2\mathbf{w}\mathbf{w}^T)^T = \mathbf{I} - 2(\mathbf{w}^T)^T \mathbf{w} = \mathbf{I} - 2\mathbf{w}\mathbf{w}^T = \mathbf{H}_1 \quad (4)$$

$$\mathbf{H}_1^T \mathbf{H}_1 = \mathbf{H}_1 \mathbf{H}_1 = (\mathbf{I} - 2\mathbf{w}\mathbf{w}^T)(\mathbf{I} - 2\mathbf{w}\mathbf{w}^T) = \mathbf{I} - 2\mathbf{w}\mathbf{w}^T - 2\mathbf{w}\mathbf{w}^T + 4\mathbf{w}\mathbf{w}^T = \mathbf{I} \quad (5)$$

Matica \mathbf{H}_1 umiestni prvý stĺpec vstupnej matice \mathbf{X} do stĺpcového vektora \mathbf{x} a to iba s jedným nenulovým prvkom. Tento nenulový prvok sa bude nachádzať na prvej pozícii. Potom pre vektor \mathbf{x} a prvý stĺpec jednotkovej matice \mathbf{I} rovný vektoru \mathbf{e}_1 platí rovnica

$$k\mathbf{e}_1 = \mathbf{H}_1\mathbf{x} \quad (6)$$

Pre konštantu k platí, že $k = \|\mathbf{x}_1\|$. Po rozpísaní rov. (6) dostaneme

$$\begin{bmatrix} k \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \mathbf{H}_1 \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix} \quad (7)$$

Vstupnú maticu \mathbf{X} označíme ako \mathbf{X}_0 a následne vykonáme jej súčin s maticou \mathbf{H}_1 . Výstupom tohto súčinu bude nová matica \mathbf{X}_1 ,

$$\mathbf{X}_1 = \mathbf{H}_1 \mathbf{X}_0 \quad (8)$$

čo je možno názorne rozpísať takto

$$\begin{pmatrix} r & r & r & r \\ 0 & x & x & x \\ 0 & x & x & x \\ 0 & x & x & x \\ 0 & x & x & x \end{pmatrix} = (\mathbf{H}_1) \begin{pmatrix} x & x & x & x \\ x & x & x & x \\ x & x & x & x \\ x & x & x & x \\ x & x & x & x \end{pmatrix} \quad (9)$$

Prvý riadok matice \mathbf{X}_1 tvorí prvý riadok hľadanej matice \mathbf{R} a prvý stĺpec obsahuje nuly pod prvým prvkom. Z matice \mathbf{X}_1 je teraz možné zostrojiť pomocnú maticu \mathbf{X}'_2 . Pomocnú maticu \mathbf{X}'_2 sme dostali tak, že sme z matice \mathbf{X}_1 odstránili 1.riadok a 1.stĺpec. Takto sme dostali pomocnú maticu \mathbf{X}'_2 , ktorej veľkosť je $(n-1) \times (n-1)$. Spôsob výberu matice \mathbf{X}'_2 môžeme vyjadriť nasledovne [6]

$$\begin{bmatrix} * & \dots & * \\ \vdots & \begin{bmatrix} * & \dots & * \\ * & \mathbf{X}'_2 & \vdots \\ * & \dots & * \end{bmatrix} & * \end{bmatrix} \quad (10)$$

Prvý stĺpec novej matice \mathbf{X}'_2 tvorí vektor \mathbf{x}_2 , ktorý obsahuje $(n-1)$ prvkov a jeho veľkosť bude nová hodnota konštanty k . Z rov. (3) vypočítame normovaný vektor \mathbf{w} a z rov. (2) Householderovu maticu \mathbf{H}'_2 veľkosti $(n-1) \times (n-1)$. Aby sme však získali maticu \mathbf{H}_2 s požadovaným rozmerom $n \times n$, tak maticu \mathbf{H}'_2 doplníme jednotkovou maticou takto

$$\mathbf{H}_2 = \begin{bmatrix} \mathbf{I} & 0 \\ 0 & \mathbf{H}'_2 \end{bmatrix} \quad (11)$$

Podobne ako v rov. (7), súčinom tejto matice \mathbf{H}_2 s maticou \mathbf{X}_1 získame maticu, v ktorej teraz už prvé dva riadky sú identické s riadkami hľadanej matice \mathbf{R} , teda

$$\mathbf{X}_2 = \mathbf{H}_2 \mathbf{X}_1 \quad (12)$$

a po názornom rozpísaní

$$\begin{pmatrix} r & r & r & r \\ 0 & r & r & r \\ 0 & 0 & x & x \\ 0 & 0 & x & x \\ 0 & 0 & x & x \end{pmatrix} = (H_2) \begin{pmatrix} r & r & r & r \\ 0 & x & x & x \\ 0 & x & x & x \\ 0 & x & x & x \\ 0 & x & x & x \end{pmatrix} \quad (13)$$

Takto postupujeme ďalej, až po $(n-1)$ krokoch dostaneme maticu \mathbf{X}_{n-1} , ktorá je identická s požadovanou maticou \mathbf{R} , pričom súčin jednotlivých Householderových matic označíme

$$\mathbf{H}_{n-1}\mathbf{H}_{n-2} \dots \mathbf{H}_1 = \mathbf{H} \quad (14)$$

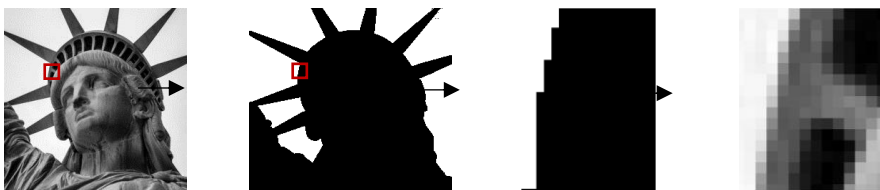
Keďže všetky matice \mathbf{H}_k sú ortogonálne, týmto aj ich súčin je tiež ortogonálny, t.j. $\mathbf{H}^T \cdot \mathbf{H} = \mathbf{I}$, zároveň platí

$$\mathbf{H}^T \mathbf{R} = \mathbf{X} \quad (15)$$

Získanú Householderovu maticu \mathbf{H} tvoria teraz ortogonálne bázové funkcie v jej riadkoch.

III. IMPLEMENTÁCIA HOUSEHOLDEROVEJ TRANSFORMÁCIE

Implementácia Householderovej transformácie na hraničné bloky objektov v obraze spracovávala ľubovoľné vstupné šedé aj farebné obrazy, na ktorých boli nejaké objekty. Pre dosiahnutie čo najlepších výsledkov bolo potrebné najprv vykonať binarizáciu obrazu, nájst' tzv. masku obrazu. Masku obrazu tvorí biele pozadie a čierny objekt a slúži na to, že po rozdelení obrazu aj masky na rovnaké bloky sa nájdu bloky, ktoré sú buď celé biele alebo celé čierne alebo čiernobiele. Čiernobiele bloky sú hraničné bloky a z nich sa na rovnakých pozíciách vyberajú hraničné bloky zo vstupného obrazu. Nájdenie binárnej masky bolo zrealizované metódou prahovania s automatickým nastavením prahu. V závislosti od typu obrazu však môže dochádzať k tomu, že v maske obrazu nájdený objekt nie je celý čierny a vykazuje nejaké nedostatky (biele miesta (škvrny) vo svojom vnútri), preto sa zrealizovala oprava týchto nedostatkov nastavením ďalších parametrov pri prahovaní. Na Obr.1 je názorná ukážka nájdenia a úpravy masky obrazu, nájdenia hraničného bloku objektu v maske a zodpovedajúceho hraničného bloku v obraze.



Obr.1 zľava: Vstupný obraz, jeho binárna maska, hraničný blok masky a hraničný blok obrazu.

Ďalej bolo potrebné zvoliť veľkosti blokov, použili sa dve veľkosti, a to 4×4 a 8×8 . Všetky pozície hraničných blokov sa ukladali do matice a následne z originálneho obrazu sa vyberali tie miesta obrazu, ktoré mali rovnakú pozíciu ako nájdené pozície pre hraničné bloky v maske obrazu. Pri hľadaní hraničných blokov sa spočítaval v maske počet hraničných blokov a počet nenulových obrazových prvkov v nich. Hraničné bloky v maske sa hľadali tak, že sa spočítal súčet všetkých prvkov v danom bloku, resp. jednotlivej matici a ak bol tento súčet väčší ako hodnota 0 a menší ako 16 pri veľkosti blokov 4×4 , tak sa daný blok vyhodnotil ako hraničný blok. Všetky nájdené hraničné bloky objektu vstupného obrazu z Obr.1 pre obe veľkosti blokov sú zobrazené na Obr.2.



Obr.2 Zobrazenie hraničných blokov pre veľkosť bloku 4×4 (vľavo) a 8×8 (vpravo).

V ďalšom kroku bolo potrebné vybrať metódu, ako sa bude vstupný obraz spracovávať. Možnosti boli nasledovné: Metóda 1 – Householderova transformácia na hraničné bloky objektu a DCT na vnútro objektu, Metóda 2 – DCT na celý obraz a Metóda 3 – DCT na hraničné bloky a vnútro objektu. Pri metódach 1 a 3 sa bloky pozadia netransformovali, iba prenášali. Miera kvality použitých metód spracovania obrazu bola vyhodnotená parametrom SNR v dB.

IV. DOSIAHNUTÉ VÝSLEDKY A ICH ZHODNOTENIE

Softvérová aplikácia na testovanie Householderovho algoritmu bola vytvorená v programovom prostredí MATLAB. Aplikácia umožňuje výber ľubovoľného vstupného obrazu, z ktorého vytvorí binárnu masku obrazu. 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 4 (inak sa obraz veľkostne upravoval). Vstupné obrazy musia byť uložené vo formátoch: jpg, bmp alebo png. Ak bol načítaný farebný obraz, následne bol prekonvertovaný na viacúrovňový obraz. Obraz sa ďalej transformoval pomocou troch uvedených metód. Výsledkom programu sú pomery SNR v [dB] celého obrazu alebo iba jeho hraničných blokov [8]. Testovanie sa uskutočnilo na skupine piatich obrazov (viď Obr. 3) veľkosti 256×256 spĺňajúcich požiadavky programu na vstupný obraz (deliteľnosť rozmeru číslom 4).



Obr.3 Vstupné testovacie obrazy.

Obrazy boli spracované 3 spôsobmi:

1. Householderova transformácia na hraničné bloky objektu a Diskrétne kosínusová transformácia na vnútro objektu – Metóda 1.
2. Diskrétne kosínusová transformácia na celý obraz – Metóda 2.
3. Diskrétne kosínusová transformácia na hraničné bloky a vnútro objektu – Metóda 3.

Dosiahnuté hodnoty SNR, ktoré sme pri simulácii navrhutej aplikácie a testovaní získali, sú pre všetky tri metódy a veľkosti blokov 4×4 uvedené v tabuľke č. 1 a pre veľkosti blokov 8×8 v tabuľke č. 2.

Tab. 1 Výsledky SNR v [dB] pre veľkosť bloku 4×4

Názov obrazu	Metóda 1		Metóda 2		Metóda 3	
	SNR [dB] pre hraničné bloky	SNR [dB] pre celý obraz	SNR [dB] pre hraničné bloky	SNR [dB] pre celý obraz	SNR [dB] pre hraničné bloky	SNR [dB] pre celý obraz
matlab	37.1	40.72	18.35	32.79	18.35	33.81
socha	35.73	24.71	17.42	22.58	17.42	22.71
znacka	36.5	41.97	27.29	15.56	15.56	28.2
lietadlo	36.24	39.69	16.78	33.17	16.78	33.45
avokado	35.05	40.89	21.23	37.32	21.23	37.42

Tab. 2 Výsledky SNR v [dB] pre veľkosť bloku 8×8

Názov obrazu	Metóda 1		Metóda 2		Metóda 3	
	SNR [dB] pre hraničné bloky	SNR [dB] pre celý obraz	SNR [dB] pre hraničné bloky	SNR [dB] pre celý obraz	SNR [dB] pre hraničné bloky	SNR [dB] pre celý obraz
matlab	28.73	39.23	21.28	32.87	21.28	33.2
socha	28.78	25.04	19.67	22.74	19.67	22.82
znacka	30.35	37.85	19.88	27.85	19.88	28.46
lietadlo	29.68	40.42	20.25	33.62	20.25	33.96
avokado	27.07	39.49	23.86	37.44	23.86	37.53

Z uvedených výsledkov pri Metóde 1 možno vidieť, že pri bloku veľkosti 4×4 bola rekonštrukcia hraničných blokov efektívnejšia ako pri veľkosti bloku 8×8. Výsledky boli oproti väčšej veľkosti bloku lepšie najmä vďaka tomu, že pri menšej veľkosti bloku hraničné bloky obsahovali menej „nadbytočných“ obrazových prvkov, a tak sa mohla Householderova transformácia aplikovať efektívnejšie. Pri veľkosti bloku 8×8 sa v blokoch, ktoré boli aplikáciou určené ako hraničné bloky, nachádzalo aj veľké množstvo prvkov, ktoré k hranici objektu v obraze nepatriли. Tieto nadbytočné prvky by mali byť súčasťou či už vnútra alebo okolia objektu, na ktoré mala byť aplikovaná DCT. Keď sa pri vyhľadávaní hraničných blokov vyberali väčšie bloky, tieto nadbytočné obrazové prvky boli súčasťou hraničných blokov, a tým sa zhoršila rekonštrukcia obrazu touto metódou.

Pri Metóde 2 sme pri rekonštrukcii hraničných blokov dosahovali takmer vo všetkých prípadoch lepšie výsledky pri blokoch veľkosti 8×8. Pri rekonštrukcii celého obrazu boli tieto hodnoty pomerne rovnaké, či už pri veľkosti blokov 4×4 alebo 8×8.

Pri Metóde 3 sme po rekonštrukcii hraničných blokov dostali rovnaké hodnoty SNR ako pri predošlej metóde. Pri oboch metódach bola na hraničné bloky aplikovaná DCT a výsledky pre hraničné bloky boli preto rovnaké ako pri Metóde 2. Pri rekonštrukcii celého obrazu sme znovu dostali pomerne rovnaké hodnoty nezávisle od veľkosti bloku.

V. ZÁVER

Článok bol zameraný na opis princípu a implementácie Householderovho algoritmu pri spracovaní hraničných blokov objektov v obraze. Z dosiahnutých výsledkov vyplynulo, že rekonštrukcia obrazu prebehla najkvalitnejšie pri použití metódy č. 1. Pri tejto metóde sme aplikovali Householderovu transformáciu na hraničné bloky objektu a na vnútro objektu v obraze

sme aplikovali DCT. Pozadie objektu sa netransformovalo. Táto metóda dosahovala oproti ostatným metódam najlepšie výsledky nezávisiac od veľkosti použitých blokov. Kvalitu rekonštrukcie obrazu sme vyhodnocovali pomocou hodnoty SNR. Kvôli porovnaniu sme odsimulovali aj ďalšie 2 metódy: aplikovanie DCT na celý obraz (označené ako Metóda 2) a aplikovanie DCT na hraničné bloky a vnútro objektu (Metóda 3). Pri porovnaní druhej a tretej metódy sme dosahovali pomerne rovnaké výsledky a v porovnaní s Metódou 1 bolo jasne vidieť značné rozdiely v zhoršenej kvalite rekonštruovaných obrazoch aj vizuálne.

Výsledky nám teda potvrdili to, že pridaním Householderovej transformácie k DCT pri spracovaní obrazov môžeme dosiahnuť lepšiu rekonštrukciu obrazu oproti použitiu klasickej DCT. Householderova transformácia skvalitní spracovanie hraničných blokov objektov, a tým sa zvýši kvalitu rekonštruovaného obrazu. Nepochádza ku skresleniu na hranici medzi objektom a pozadím ako pri klasickej DCT.

Na koniec záveru patrí poďakovanie Ing. Dominikovi Vudmaskovi, ktorý sa podieľal na riešení úloh z tejto problematiky.

LITERATÚRA

- [1] ITU-T Recommendation H.264, 03/2005.
- [2] Li, Y. a kol.: Real structure-preserving algorithms of Householder based transformations for quaternion matrices. [Online] Elsevier, 15.10.2016. Dostupné z <https://www.sciencedirect.com/science/article/pii/S0377042716301583>
- [3] Salam, A. – El Farouk, A. – Al-Aidarous, E.: Symplectic Householder transformations for a QR-like decomposition, a geometric and algebraic approaches [Online] Elsevier, 05.2008. Dostupné z <https://www.sciencedirect.com/science/article/pii/S0377042707001537>
- [4] Golub, G. H. – Van Loan, Ch. F.: Matrix computations [Online] JHU Press, 15.02.2013. Dostupné z https://books.google.sk/books?id=5U-l8U3P-VUC&printsec=frontcover&hl=sk&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false
- [5] Salam, A. – El Farouk, A. – Al-Aidarous, E.: Optimal symplectic Householder transformations for SR decomposition [Online] Elsevier, 01.09.2008. Dostupné z <https://www.sciencedirect.com/science/article/pii/S002437950800116X>
- [6] Mihalík, J. – Gladišová, I.: Diskrétne ortogonálne transformácie vizuálnych objektov (Algoritmy a riešené príklady). Dostupné z https://www.researchgate.net/publication/335395943_DISKRETNE_ORTOGONALNE_TRANSFORMACIE_VIZUALNYCH_OBJEKTOV_Algoritmy_a_riesene_priklady
- [7] Rotella, F. – Zambettakis, I.: Block Householder transformation for parallel QR factorization [Online] Elsevier, 05.1999. Dostupné z <https://www.sciencedirect.com/science/article/pii/S0893965999000282>
- [8] Hossack, W.: Noise in Images [Online]. The University of Edinburgh, Dostupné z <https://www2.ph.ed.ac.uk/~wjh/teaching/dia/documents/noise.pdf>

A measurement method of basic ferromagnetic material properties

¹Šimon GANS, ²Ján MOLNÁR

^{1,2} Department of Theoretical and Industrial Engineering, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Slovak Republic

¹simon.gans@tuke.sk, ²jan.molnar@tuke.sk

Abstract — This paper addresses the theoretical basis and practical realization of a simple measurement method of material properties of ferromagnetic materials. When choosing a suitable magnetic material for a specific use case whether it is for example for the construction of electric motors, transformers or permanent magnets, different material properties must be considered. Some materials are easily magnetizable, some are magnetized with a lot of difficulty. After being magnetized into saturation, some materials are left with a lot of remanent magnetic flux, some less so. Measuring the so-called BH curve of a specific material can provide a lot of information about the best practical application of the material. Special automatized measuring equipment of these curves can be expensive, which is the reason why the paper is focusing on an easy and cheap method to execute this type of measurement.

Keywords — BH curve, ferromagnetic materials, magnetization, measurement

I. INTRODUCTION

The automated industrial world of today is based on electric motors. In the long-range distribution of electrical energy transformers are employed to change the voltage levels for more efficient energy transfer. In some types of motors and generators permanent magnets are also employed. Ferrimagnetic materials are sometimes used instead of ferromagnetic materials because of their low electrical conductivity, but they have their own drawbacks in other aspects. Many factors must be considered when assigning a specific use case for a material[1].

The most important material characteristics which can be extracted from the BH curve can be observed in the figure below (Fig. 1). The intersections of the curve on the y-axis are the remanent magnetic flux densities B_R , the intersection points on the x-axis are the coercive forces H_C , the points at which the material magnetically saturates are known as the magnetic flux density at saturation B_S . Also the area inscribed by the hysteresis loop represents the heat energy density generated by the cyclic magnetization of the material in one cycle due to hysteresis losses[1],[2].

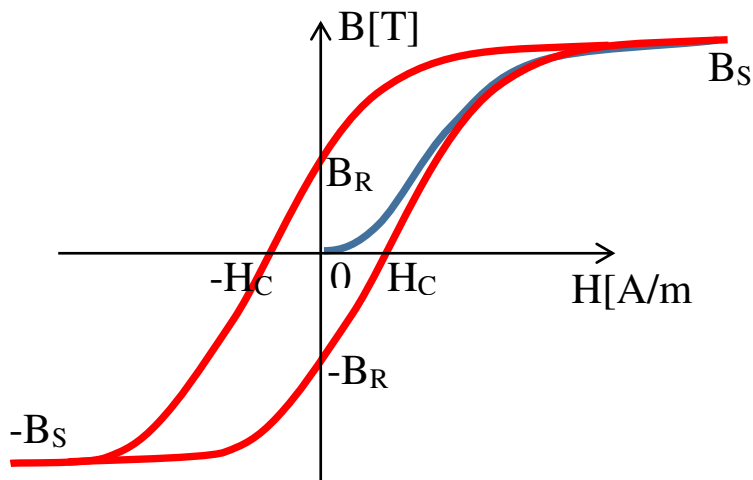


Fig. 1: Example of a BH curve of a material. BH curves of real materials are hysteretic - they form closed loops when magnetized periodically. When the material at the start is fully demagnetized, it will be firstly magnetized along the initial magnetization curve (blue curve). After that, if periodically magnetized with an AC field, the magnetic flux B will not go back to zero, but will run along the red hysteretic curve.

The hysteresis power loss can also be approximated by the following equation (1), where η is the Steinmetz coefficient of the material, B_{max} is the highest value of magnetic flux density obtained in the material during magnetization, f is the magnetizing frequency, n is the Steinmetz exponent and V is the volume of the material[2].

$$P_{hys} = \eta \cdot B_{max}^n \cdot f \cdot V \quad (1)$$

The hysteresis loss grows with frequency and with the value of B_{max} . The lesser maximum flux density the material experiences, the smaller the area of the hysteresis curve becomes and so the power losses are lower as well. These lower flux hysteresis loops are known as minor loops and examples of them are depicted in the picture below (Fig. 2)[3].

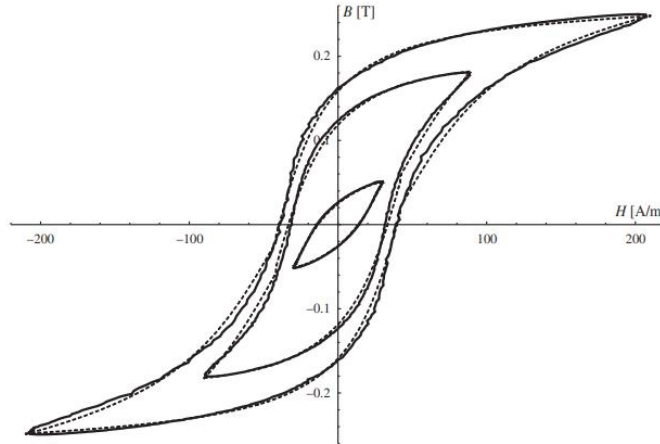


Fig. 2: Measured (solid lines) minor hysteresis loops of a MnZn ferrite material. The dashed lines represent the minor loop predictions estimated from the mathematical model of ferromagnetic materials developed in study[3].

II. TYPES OF MATERIALS WITH STRONG MAGNETIC BEHAVIOR

In general, strongly magnetic materials (ferromagnetic and ferrimagnetic materials) can be divided into two groups[4]:

- Soft magnetic materials – these materials are easily magnetized and demagnetized. They are best used in cases where fast switching speeds of AC currents are present in the magnetizing coils. The amount of flux retained by the material after the magnetizing field vanishes is low. Their permeability is high (permeability is the steepness or the instantaneous rate of change of the initial magnetization curve (blue graph in Fig. 1)). Their hysteresis losses are small, and such materials are used for example for transformer or inductor core construction[4].
- Hard magnetic materials – these materials are difficult to magnetize and demagnetize. The magnetic flux density which remains in the material after the magnetization process is high. The Curie point (the temperature at which the ferromagnetic material loses its magnetic properties and becomes paramagnetic) is usually high too. The hysteresis loop of such a material has a large area which makes the hysteresis losses quite significant especially at higher frequencies[4].

A table is shown below (Tab. 1) in which the most important material characteristics and their typical numerical values are listed. Magnetic hardness has no correlation with mechanical hardness[4],[5].

Tab. 1: Basic comparison between soft and hard magnetic materials

	Soft magnetic materials	Hard magnetic materials
Coercive force H_C	Small values (as low as 1 A/m)	Large values (as big as 400 kA/m)
Hysteresis loop	Small area and steep curves	Large area and slowly rising curves
Hysteresis loss	Small (as low as mW per kg)	Large (as big as kW per kg)
Relative permeability	High (as large as 10^6)	Low (as small as 10^1)
Saturation flux B_S	High (as large as 2 T)	
Remanent flux B_R	Lower (usually < 1 T)	Higher (usually > 1 T)
Uses	Transformers, inductors, motors	Permanent magnets, relays, motors

III. THEORETICAL BACKGROUND OF MAGNETIC MEASUREMENTS

The relationship between the magnetic flux density B and magnetic field intensity H in vacuum is shown in equation (2), where μ_0 is the permeability of free space[6].

$$B = \mu_0 H \quad (2)$$

When a magnetic field H exists inside of a material in space, then the magnetization of the material M adds to the field H which magnetizes the material, making the magnetic flux density B much larger. This property is especially pronounced in ferromagnetic and to a somewhat lesser amount in ferrimagnetic materials and it is expressed by equation (3)[6].

$$B = \mu_0(H + M) \quad (3)$$

Alternatively, equation (3) can be written in a more compact form (4), in which a new quantity, the relative permeability μ_r is introduced. It expresses the influence of the material on the size of the magnetic flux density B at a specific magnetic field strength H [6].

$$B = \mu_0 \mu_r H \quad (4)$$

B and H must be measured at equal time points and then plotted against each other to plot the BH curve. Coils of wire are wound around the ferromagnetic sample to create the magnetic field intensity H inside of the sample and to sense the magnetic flux density B flowing through the sample[6].

A. Magnetizing and sensing coils

The figure below (Fig. 3) shows that when current is flowing through the windings of a solenoid, then the magnetic field H created in a solenoid is expressed by formula (5), where N is the number of turns of the coil, I is the current flowing through the coil, d is the diameter of the circular solenoid cross-section and l is the length of the coil[7].

$$H = \frac{N \cdot I}{l \sqrt{1 + \left(\frac{d}{l}\right)^2}} \quad (5)$$

The magnetic field on the axis of the solenoid varies slightly when moving away from the geometric center of the coil and in some applications, this is solved by increasing the winding density near the ends of the coil, but in technical applications this can be to a certain degree omitted. When the ratio between the coil cross-section diameter and the coil length is small enough, formula (5) can be approximated by formula (6). It is evident, that to measure the magnetic field intensity, the coil current is measured[7].

$$H = \frac{N \cdot I}{l} \quad (6)$$

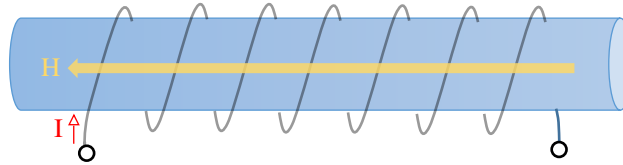


Fig. 3: The magnetic field generated inside a solenoid, which has a current flowing through it. The current creates a homogenous magnetic field inside of the coil of an intensity which is directly proportional to the flowing current.

When a homogeneously wound solenoid is considered, then the voltage generated on its ends is expressed by formula (7), where $u(t)$ is the generated voltage, S is the area of the coil windings (assuming every turn is identical) and B is the magnetic flux density flowing through the coil. It is evident that only a time-varying flux density induces voltage in the coil[6].

$$u(t) = -N \cdot S \cdot \frac{dB}{dt} \quad (7)$$

When the magnetic flux density is expressed from equation (7) by integration, equation (8) is created, where B_0 is the magnetic flux intensity at time $t = 0$ [8].

$$B(t) = -\frac{1}{N \cdot S} \cdot \int_0^t u(t) dt + B_0 \quad (8)$$

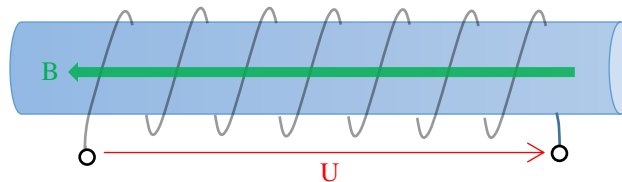


Fig. 4: When a changing magnetic flux density B is flowing perpendicular to the surfaces that are bounded by the coil turns (so flowing parallel to the coil axis), then voltage is generated between the open ends of the coil.

The magnetic flux is proportional to the time integral of the voltage generated on the coil ends. To create a time varying magnetic flux density B a time varying magnetic field intensity H must

be created by the magnetizing coil, which means that the magnetizing current must be time varying as well[6].

B. Shape of the sample

When a ferromagnetic sample has open ends in the direction of magnetization like the cylinder in Fig. 3, the open ends generate a demagnetizing field, which has opposite effects than the magnetizing field created by the coil inside of the sample. This magnetic pole field strength is expressed by equation (9), where N_Z is the demagnetizing factor dependent on the geometry. There is a negative sign in the formula which emphasizes that the demagnetizing field is acting against the field strength[7].

$$H_p = -N_Z \cdot (\mu_r - 1) \cdot H \quad (9)$$

This field changes the shape of the BH curve and shifts its values so the measurements must be corrected afterwards. The real field intensity values H_{sample} that the sample experiences are given by equation (10). If the sample has a geometry with a demagnetizing factor equal to 0, no correction is required. Toroid shape cores have such a property in the direction of the toroid core length. Such a toroid core is depicted in the figure below (Fig. 5)[7].

$$H_{sample} = H \cdot (1 - N_Z(\mu_r - 1)) \quad (10)$$

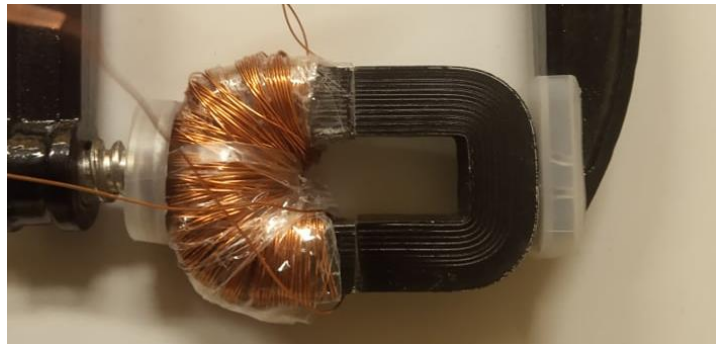


Fig. 5: A toroid shaped ferromagnetic core consisting of two equal pieces pressed against each other for a lower (but not zero) magnetic resistance between the two halves. The coil length in this case is represented by the length of the ferromagnetic core – the average field length, since we can assume that all the field lines created by the coil are staying in the ferromagnetic core. The magnetizing coil is wound over the sensing coil.

IV. MEASUREMENT SETUP

The measurement setup can be seen on the pictures below (Fig. 6). On that picture the red rectangle is the energy source for the magnetization of the sample. It consists of a regulated AC sinusoidal voltage source with a controllable output RMS voltage from 0 to 250 V with 1 V resolution. It is connected to a step-down transformer with a transformation ratio of 10:1 to make the supply voltage resolution approximately 100 mV. The sinusoidal voltage will create a current which magnetizes the sample.

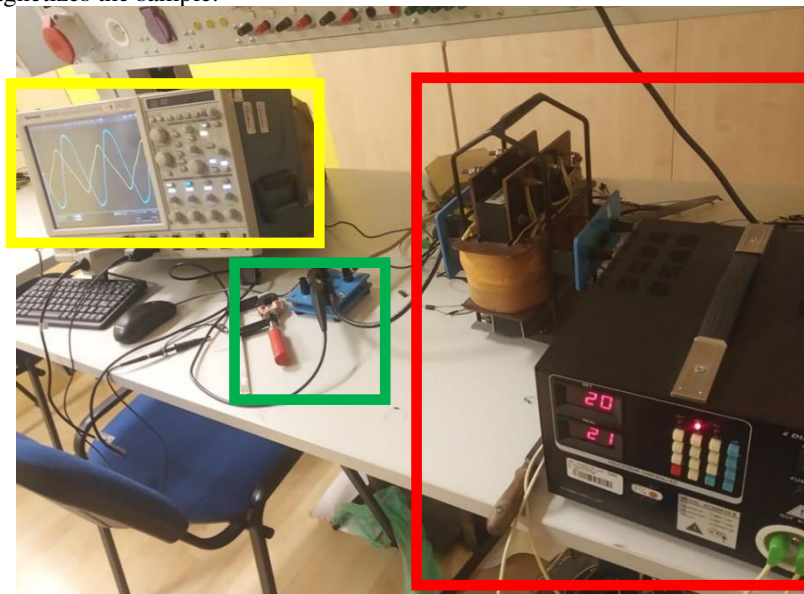


Fig. 6: The practical realization of the measurement setup

The green rectangle depicts the measured sample, which is more closely viewed in the figure below (Fig. 7) with the output voltage (sensing coil) and input current (magnetizing coil) measurements. The probes are connected to the oscilloscope (yellow rectangle) where the waveforms are depicted on the figure below (Fig. 8).

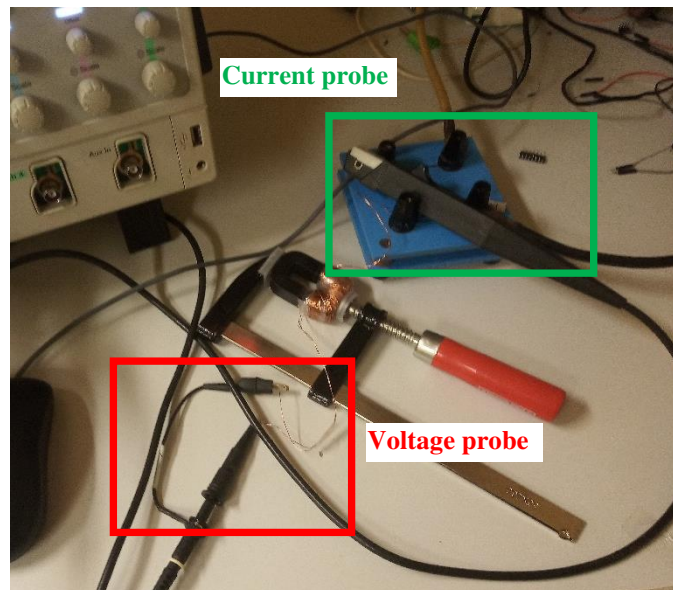


Fig. 7: Close up image of the measured sample and the measurement probes. The clamps holding the toroid core were lightly tightened only to make sure that the toroid halves cross-sections perfectly overlap and are kept in contact. The amount of mechanical stress introduced into the core was very insignificant and can be neglected, even though soft ferromagnetic materials are sensitive to mechanical stress.

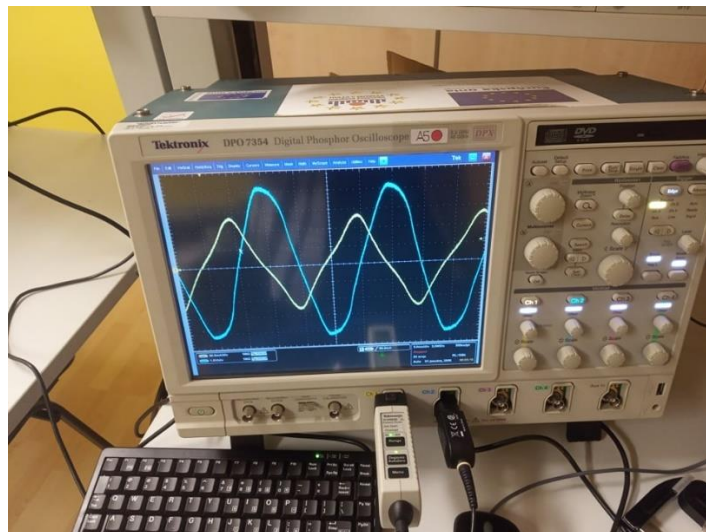


Fig. 8: The oscilloscope measurements of the toroidal ferromagnetic core. The blue waveform represents the output voltage and the yellow one the input current. Multiple measurements have been taken at different saturation levels.

After measuring the magnetic core length (12,9 cm), core cross-section area (1,25 cm²), putting in the number of primary (80) and secondary turns (80) of the windings and measuring the core's output voltage and current, the output voltage was integrated numerically on the oscilloscope and multiplied by the constant in equation (8) to get the values of the magnetic flux density B inside the material. The current was multiplied by the constants in equation (6) to obtain the values of the magnetic field intensity in the material H . The data was then exported to a USB drive plugged in the oscilloscope during the measurement and was then exported into excel for graphing. The BH curve of the material can be seen below (Fig. 9).

From the major loop the remanent magnetic flux B_R is approximately 0,6 T, the coercive force is 25 A/m (which is also affected by the air gap between the core halves). Some soft magnetic materials have even lower values of the coercive force H_C . The saturation magnetization for this material is around 1,85 T. The measured data of B and H were smoothed out by using a moving average with a window-width of 16, because the noise made the measured BH curves jittery and unclear, which was corrected by this post-processing step. The graph was also shifted downwards a bit, subtracting the average value of the magnetic flux density B throughout the measurement

interval because when the integration process started at time $t=0$, there was some magnetic flux present in the core B_0 which was unknown. The correction consists of only shifting the BH curve graph so that the values of $+B_R$, $+H_C$ and $+B_S$ and $-B_R$, $-H_C$ and $-B_S$ are symmetrical (their absolute values should be theoretically identical). Sometimes when DC noise (or slowly time-varying noise) is superimposed to the measured voltage, the output from the integration process yields a steadily increasing part added to the AC sinusoidal output voltage which needs to be compensated for, otherwise the BH curves has “ruffled” edges (multiple saturation values) since in every cycle the saturation flux shifts due to this steadily increasing/decreasing noise flux. This error can be easily detected by methods of linear interpolation of the measured data (B vs. time and H vs. time) and compensated for. The area of the BH curve will be computed in future work.

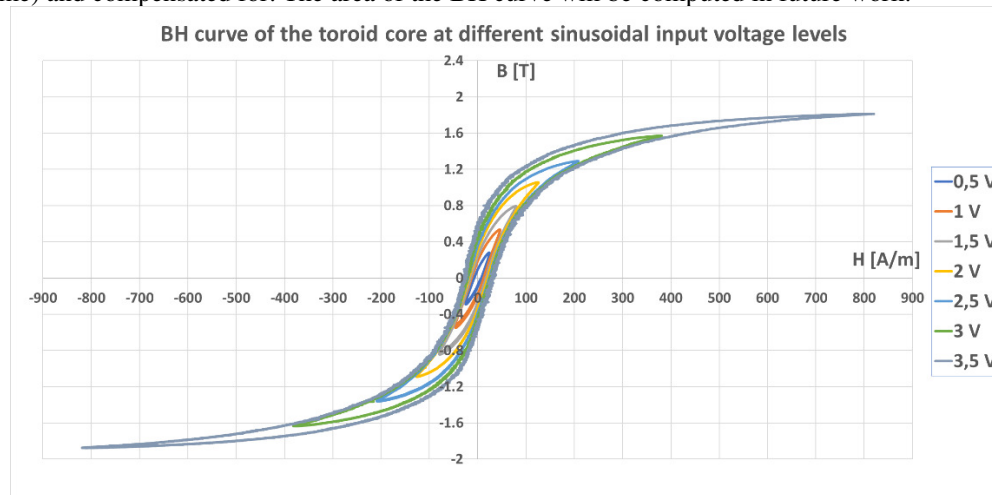


Fig. 9: The measured minor and major BH curves of the material. The lesser the supply voltage is, the lesser current flows through the magnetizing winding and so the BH curve reaches smaller maximum magnetic fluxes which results in smaller BH loops.

V. CONCLUSION

A fast and easy way to measure magnetization curves has been presented in this paper. Further work however is needed to make this process more automated. Not needing to export the data from an oscilloscope would make the measurement process faster for example by employing a microcontroller system that would measure the current and voltage and integrate the waveforms in real-time and send the data to a database via Wi-Fi from which the graphs would be drawn. Creating a script that will compute the area of the BH curve, for hysteresis loss estimation will also be the subject of future work.

ACKNOWLEDGMENT

The authors thank for the financial support from the GRANT FEI 2022 project for the grant specified by the index FEI-2022-82 which helped in the research.

REFERENCES

- [1] WatElectronics, *What are Ferromagnetic Materials: Magnetization & Their Applications*, 2020. Available on the internet at: <https://www.watelectronics.com/what-are-ferromagnetic-materials-magnetization-their-applications/>
- [2] D. Collins, *Hysteresis loss and eddy current loss: What's the difference?*, 2018. Available on the internet at: <https://www.motioncontroltips.com/hysteresis-loss/>
- [3] A. M. Milovanovic, B. M. Koprivica, *Mathematical Model of Major Hysteresis Loop and Transient Magnetizations*, 2015, *Electromagnetism*, 35:3, 155-166, DOI:10.1080/02726343.2015.1005202. Available on the internet at: https://www.researchgate.net/publication/273143832_Mathematical_Model_of_Major_Hysteresis_Loop_and_Transient_Magnetizations
- [4] ElectricalVoice, *Soft and Hard Magnetic Materials*, 2017. Available on the internet at: <https://electricalvoice.com/soft-and-hard-magnetic-materials/>
- [5] H. Reinboth, *Vlastnosti a použití magnetických materiálů*, SNTL-Nakladatelství technické literatury, Praha, 1975, 04-535-75, pages 35-38, translated 3rd edition
- [6] J. F. Cochran, B. Heinrich, 6.2: *B-H Curves*, 2022, Simon Fraser University. Available on the internet at: [https://phys.libretexts.org/Bookshelves/Electricity_and_Magnetism/Book%3A_Applications_of_Maxwells_Equations_\(Cochran_and_Heinrich\)/06%3A_Ferromagnetism/6.02%3A_B-H_Curves](https://phys.libretexts.org/Bookshelves/Electricity_and_Magnetism/Book%3A_Applications_of_Maxwells_Equations_(Cochran_and_Heinrich)/06%3A_Ferromagnetism/6.02%3A_B-H_Curves)
- [7] V. Hajko, L. Potocký, *Fyzika mangetických javov*, 1973, Rektorát Univerzity P. J. Šafárika v Košiciach, ES-UPJŠ 10/64/73. Pages 185-189.
- [8] M. J. Dapino, R. C. Smith, F. T. Calkins, A. B. Flatau, *A Coupled Magnetomechanical Model for Magnetostrictive Transducers and its Application to Villari-effect Sensors*, 2002, *Journal of Intelligent Material Systems and Structures*, DOI: 10.1177/1045389X02013011005. Available on the internet at: https://www.researchgate.net/publication/240600640_A_Coupled_Magnetomechanical_Model_for_Magnetostrictive_Transducers_and_its_Application_to_Villari-Effect_Sensors/figures?lo=1

Princípy syntetizácie lekárskych správ

¹Lucia GOJDIČOVÁ, ²Michal KOLÁRIK

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

¹lucia.gojdicova@student.tuke.sk, ²michal.kolarik@tuke.sk

Abstrakt — Hlavným cieľom tohto článku je stručne predstaviť syntetizáciu údajov v medicínskej oblasti. Predstaviť metódy pre syntetizáciu a experimentálne overiť vybranú metódu pre generovanie syntetických údajov. Syntetické údaje boli vytvorené pomocou modelu CTGAN a výsledky boli overené vybranými metrikami ako aj vizuálne.

Kľúčové slová — syntetizácia, GAN, CTGAN, generovanie syntetických údajov

I. ÚVOD

Údaje z medicínskej oblasti sa verejne publikovať nemôžu, pretože podliehajú zákonu, lekárskej tajomstvu, alebo sa ľudia boja zneužitia osobných údajov. Na to aby bolo možné analyzovať takéto dáta je nutné ich odosobniť. Anonymizáciu údajov môžeme využiť pre anonymizovanie osobných údajov, kde pri tejto metóde dochádza k odstraňovaniu alebo nahrádzaniu údajov.

Ďalším riešením ochrany súkromia je syntetické generovanie údajov za účelom vytvorenia náhradných syntetických údajov. Takéto syntetické údaje môžeme považovať ako doplnenie k štandardným anonymizačným postupom. Tento prístup nám umožní vytvoriť dataset veľmi podobný originálnemu datasetu avšak plne anonymizovaný, taký, ktorý môžeme následne využiť pre dátovú analýzu alebo publikovanie. Na takéto generovanie syntetických údajov sme v rámci tejto práce popísali jednotlivé metódy, vďaka ktorým môžeme syntetické údaje získať. Medzi metódy sme zaradili rozhodovacie stromy, Bayesovské siete, metódu copulas a metódu GAN a jej odvodené architektúry napríklad CTGAN a MedGAN.

V práci sme sa zamerali na vytvorenie syntetických údajov pomocou metódy GAN a jej modelu CTGAN. Pôvodný súbor údajov sme použili na tréning generatívneho modelu pomocou strojového učenia, a následne sme tento model použili na vytvorenie realistických, no umelých údajov, ktoré však majú rovnaké štatistické vlastnosti ako pôvodný súbor. Popísali a použili sme rôzne metriky a vizualizácie na zistenie kvality, spoľahlivosti a podobnosti, ktorých výsledky spomenieme.

II. ANONYMIZÁCIA A SYNTETIZÁCIA ÚDAJOV

Hlavnou úlohou anonymizácie údajov je pripraviť osobné údaje tak, aby nedošlo ku zisteniu identity. Anonymizáciou údajov dochádza ku odstraňovaniu alebo nahrádzaniu osobných údajov pomocou rôznych metód, ktoré sú na to určené. Môžeme spomenúť napríklad metódu náhrady, kde pôvodné údaje sú nahradzované z externého zdroja. Ďalej metóda nahrádzania hodnotami NULL a NA, ktoré nahradia citlivé údaje, alebo metóda vymazanie, kde sa odstraňujú stĺpce s takými osobnými údajmi. Takéto metódy anonymizácie sú jedným z riešení ochrany údajov.

Druhým riešením je syntetizácia údajov, ktorá sa dostala do popredia so strojovým učením. Syntetické údaje predstavujú umelo vytvorené údaje a celková syntetizácia je považovaná za kľúčovú technológiu, ktorá pomáha zlepšovať súkromie [1]. Dôležitými vlastnosťami týchto syntetických údajov je, že majú schopnosť napodobniť dôležité vlastnosti skutočných údajov bez toho, aby bolo súkromie odhalené. Taktiež na rozdiel od iných metód zostávajú takéto údaje odolné voči opätovnej identifikácii. K opätovnej identifikácii dôjde, keď sa útočníkovi pomocou pomocných informácií podarí správne priradiť identitu konkrétnej osoby k určitému záznamu v uvoľnenej anonymizovanej množine údajov. Pri syntetických údajoch je nový dataset generovaný z pôvodného, ktorý sa na neho iba podobá a má rovnaké štatistické vlastnosti. Pri anonymizácii sa pracuje s pôvodným datasetom, kde prebieha spomínané nahrádzanie a odstraňovanie.

III. METÓDY A METRIKY

Na vytváranie syntetických údajov môžeme použiť metódy, ako sú rozhodovacie stromy, bayesovské siete, copulas a metódu GAN. Rozhodovacie stromy sa primárne používajú na klasifikáciu alebo predikciu, kde sa stretávame s vysokou presnosťou, jednoduchou reprezentáciou a používajú sa v rôznych medicínskych oblastiach. Bayesovské siete predstavujú pravdepodobnostné modely založené na acyklicky orientovaných grafoch. Tieto Bayesovské siete si môžeme predstaviť ako sieť vytvorenú z uzlov a hrán, ktoré predstavujú modelové premenné a vzťahy medzi nimi. Metóda copulas plní dôležitú úlohu pri závislosti medzi dvoma alebo viacerými premennými. V rámci strojového učenia slúži na pomoc pri generovaní syntetických tabuľkových údajov. Metóda GAN reprezentuje dve protichodné siete, ktoré stoja proti sebe. Je typom architektúry neurónových sietí. Obsahuje dve takéto neurónové siete, konkrétne generátor a diskriminátor. Generátor generuje nové údaje, zatiaľ čo diskriminátor sa pokúša správne klasifikovať skutočné a falošné údaje. Cieľom generátora je oklamať diskriminátor. Metódu GAN môžeme využiť pre štruktúrované a neštruktúrované údaje (obrázky a text).

V práci sme sa zamerali na metódu GAN [2] a vybrali ju ako najvhodnejšiu metódu na generovanie syntetických údajov. Môžeme spomenúť viacero architektúr, ktoré sú odvodené z GAN, ako napríklad MedGAN [3], CTGAN [4], TableGAN [5] a CopulaGAN [2], ktoré pracujú s tabuľkovými údajmi. V práci sme si pre generovanie tabuľkových údajov vybrali model CTGAN, ktorý priniesol kvalitné výsledky. Model CTGAN predstavuje metódu na modelovanie tabuľkových údajov. Využíva sa na generovanie nových syntetických údajov, pričom sa zachovávajú rovnaké štatistické vlastnosti s pôvodnými údajmi. Tento model bol prezentovaný na konferencii NeurIPS 2020.

Na vyhodnotenie výsledkov modelu sa používajú rôzne metriky ako štatistické testy a kombinovaná metrika. Pri štatistických testoch poznáme dva takéto testy, KSTest [2] a CStest [2]. S kategorickými atribútmi, ktoré sa nachádzajú v datasete pracuje CStest a KSTest hodnotí kvalitu pri numerických atribútoch. KSTest porovnávaná stĺpce a výsledkom je maximálna vzdialenosť medzi pozorovanými a očakávanými hodnotami, pri CSteste je výsledkom p hodnota. Tieto hodnoty oboch testov sú v rozpätí 0 až 1. Kombinovaná metrika¹ sa skladá z viacerých metrík a porovnávaná pôvodné a syntetické údaje. Výsledkom je hodnota taktiež v rozpätí 0 až 1, ktorá nám hovorí ako veľmi sa tieto údaje podobajú. Na porovnanie údajov rozhodovacím stromom sme použili metriky ako úspešnosť, presnosť, návratnosť a F1 – skóre.

IV. VYTVÁRANIE SYNTETICKÝCH DATASETOV

Na vytváranie syntetických údajov sme použili metódu CTGAN, ktorá pracuje s tabuľkovými údajmi. Zvolili sme programovací jazyk Python, pretože má veľa knižníc, ktoré sa používajú práve na syntetizáciu údajov v tomto jazyku, a taktiež má vynikajúcu podporu. Z ponuky knižníc sme vybrali knižnicu SDV (Synthetic Data Vault). Táto knižnica pomocou strojového učenia slúži ako nástroj na modelovanie datasetov.

Na generovanie nových syntetických údajov sme zvolili tri veľkostne odlišné datasety z oblasti medicíny. Datasety pozostávajú z údajov o pacientoch s hepatitídou C, rakovinou a kardiovaskulárnymi chorobami. Pri najväčšom datasete sme použili vzorku 10 000 záznamov z pôvodných 70 000 vzhľadom na výpočtovo náročné modelovanie a generovanie syntetických údajov. Generovanie pomocou modelu CTGAN prebiehalo na každom datasete samostatne.

Generovanie pozostávalo z týchto krokov:

1. Vytvorili sme inštanciu modelu.
2. Zavolali sme metódu fit, kde sme odovzdali naše pôvodné údaje vo forme tabuľky.
3. Následne prebiehalo modelovanie.
4. Ukončenie modelovania.
5. Generovali sme nové syntetické údaje zavolaním sample metódy.
6. Zvolili sme si vlastný počet riadkov, ktoré budú obsahovať syntetické údaje.

Výsledkom generovania je nová tabuľka so syntetickými údajmi, ktoré majú podobné štatistické vlastnosti ako pôvodné údaje.

V. VÝSLEDKY, ZISTENIA A PRÍNOSY

Model CTGAN obsahuje rôzne hyperparametre, ktoré ovplyvňujú kvalitu údajov. Na to aby boli nové syntetické údaje čo najviac kvalitné, menili sme hyperparametre epochs a batch_size. Vykonal sme niekoľko nových generovaní, kde sme nastavovali hodnoty na 100, 500 a 1000, a opäť kontrolovali kvalitu a podobnosť nových syntetických údajov. Po niekoľkých pokusoch, kde sme využívali Kombinovanú metriku a štatistické testy, sme zistili optimálne nastavenie

¹ Kombinovaná metrika [Online]: https://sdv.dev/SDV/user_guides/evaluation/evaluation_framework.html

parametrov na hodnoty 1000 pre epochs a 500 pre batch_size. Výsledok je možné vidieť v tabuľke Tabuľka 1.

Tabuľka 1.
Kvalita vytvorených syntetických datasetov

Datsety	KSTest	CSTest	Kombinovaná metrika
Hepatitída C	0.78	0.99	0.57
Rakovina	0.87	0.99	0.48
Kardiovaskulárne choroby	0.92	-	0.67

Pred samotným generovaním syntetických údajov sme pomocou funkcie *duplicated()* hľadali duplicitné záznamy v pôvodných údajoch. Takéto záznamy sa nám objavili v dvoch datasetoch, konkrétne v strednom a najväčšom datasete. Hľadaním duplicitných záznamov sme pokračovali. Postupne sme odoberali po jednom jednotlivé atribúty v každom datasete samostatne a hľadali duplikáty. Pri najmenšom datasete, ktorý nemal duplicitné záznamy na úrovni celého datasetu, sme v prípade odobratia atribúty *Age* zistili duplicitné záznamy. Pri odoberaní ďalších atribútov sme ďalšie duplikáty nenašli. Pri strednom datasete, pri odoberaní jednotlivých atribútov, sme dostávali číslo 848, ktoré sa zhodovalo s hodnotou duplicitných záznamov na úrovni celého datasetu. Pri najväčšom datasete sme si všimli zaujímavosť pri atribúte *Age*, kde po jeho odobratí sme získali hodnotu 17597 duplicitných záznamov zo 70 000 záznamov, príčinou bola štruktúra tabuľky, kde všetky ostatné atribúty boli kategorické s obmedzeným množstvom položiek.

Po vygenerovaní nových syntetických údajov sme taktiež použili funkciu *duplicated()* na hľadanie duplicitných záznamov. Hľadaním duplicitných záznamov sme zisťovali koľko záznamov z pôvodného datasetu sa objavuje v syntetickom datasete. Duplicitné záznamy sme hľadali v pôvodných a syntetických údajoch, a zistili sme, že sa takéto údaje nachádzajú iba v pôvodných údajoch. Taktiež sme pôvodné a syntetické údaje spojili spolu, kde sa nám potvrdilo tvrdenie, že duplicitné záznamy sa nachádzajú iba v pôvodných údajoch. Mohli sme konštatovať, že aj napriek výskytu duplicitných záznamov v pôvodných údajoch, v syntetických sa takéto údaje nenachádzali.

Spomenutí model rozhodovacieho stromu sme využili na porovnanie pôvodných a syntetických údajov, kde sme vykonali tri experimenty. Vykonali sme experimenty na všetkých troch pôvodných datasetoch a všetkých troch syntetických. Ako prvé sme trénovali a testovali modely na pôvodných údajoch, ďalej sme trénovali a testovali modely na syntetických údajoch, a ako posledné v poradí sme modely natrénované na syntetických dátach testovali na pôvodných údajoch.

Na vyhodnotenie sme použili úspešnosť, presnosť, návratnosť a F1 – skóre. Taktiež sme použili konfúznú maticu na porovnanie predikovaných hodnôt, získaných pomocou rozhodovacieho stromu a reálnych hodnôt z testovej množiny. Rozhodovací strom, ktorý sme použili na pôvodné údaje dosahoval vo všetkých troch experimentoch výborné alebo veľmi dobré výsledky. Napríklad pri metrike F1 – skóre sme v jednom z experimentov dosiahli hodnotu jedna, čiže sto percent. Pri syntetických údajoch pozorujeme podobné výsledky, avšak pri testovaní reálnych údajov na syntetických pozorujeme v niektorých prípadoch zníženie výsledkov. Tieto výsledky môžeme odôvodniť menším množstvom záznamov v syntetickej množine, vzhľadom na pôvodnú množinu, čiže model sa to nevie naučiť. Pri najväčšom datasete, kde sme syntetickú množinu testovali na pôvodnej množine si môžeme všimnúť podobné výsledky presnosti, mohlo ísť o to, že v datasete sa nachádzajú iba numerické atribúty. Výsledky experimentov sú dostupné v tabuľke Tabuľka 2.

Tabuľka 2.
Kvalita vytvorených syntetických datasetov

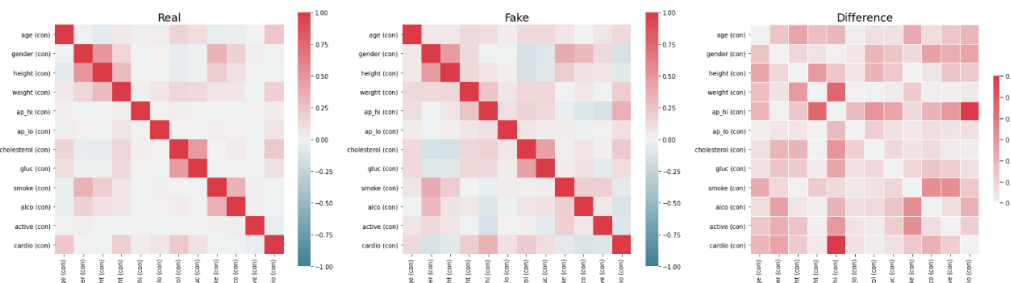
Datsety	správnosť	návratnosť	F1-miera
Pôvodný dataset			
Hepatitída C	0.90	0.90	0.90
Rakovina	1.00	1.00	1.00
Kardiovaskulárne choroby	0.73	0.73	0.73
Syntetický dataset			
Hepatitída C	0.49	0.59	0.53
Rakovina	0.92	0.91	0.91
Kardiovaskulárne choroby	0.72	0.72	0.72

Model trévaný na syntetickom datasete a testovaný na pôvodnom datasete			
Hepatitída C	0.83	0.78	0.80
Rakovina	0.78	0.79	0.78
Kardiovaskulárne choroby	0.70	0.70	0.70

V ďalšom kroku sme a vykresľovali natrévané rozhodovacie stromy v grafickej podobe a porovnávali. Sledovali sme, aké zmeny sa objavili v modeloch trévaných na syntetických dátach v porovnaní s modelmi trévaným na pôvodných údajoch. Pri najmenšom datasete sme si všimli, že obidva rozhodovacie stromy sa na začiatku rozhodujú rovnako, avšak pri syntetických údajoch následne dochádzalo k viacerým rozvetveniam. Pri strednom datasete sme si všimli úplnú rozdielnosť výsledných stromov. Najviac podobné grafické zobrazenie stromu sme pozorovali pri najväčšom datasete, v ktorom sa nachádzali iba numerické atribúty.

Na grafické porovnanie pôvodných a syntetických údajov sme použili najprv knižnicu Table Evaluator a následne funkciu *visual_evaluation()*. Pomocou tejto funkcie sme zobrali absolútny logaritmický priemer a strednú odchýlku numerických údajov pôvodných a syntetických údajov pre jednotlivé atribúty, ďalej kumulatívny súčet cieľového atribútu a nakoniec korelačné tabuľky, ktoré zobrazovali pôvodné a syntetické údaje a taktiež aj rozdiel medzi nimi.

Pri grafických vizualizáciách sme dosahovali veľkú podobnosť medzi pôvodnými a syntetickými údajmi. Najväčšiu podobnosť sme dosahovali pri strednom a najväčšom datasete. Naopak najmenej sa podobali údaje z najmenšieho datasetu. Práve týmito vizualizáciami môžeme potvrdiť to, že syntetické údaje sú vhodné ako náhrada za tie pôvodné.



Obrázok 1. Korelácia atribútov v pôvodnom datasete (Real) kardiovaskulárnych chorôb, syntetickom (Fake) datasete a ich vzájomnom rozdiel (Difference).

VI. ZÁVER

V uvedenej práci bola predstavená anonymizácia a syntetizácia údajov. Taktiež sme popisovali jednotlivé metódy syntetizácie pre vytváranie syntetických údajov. Následne sme si vybrali metódu GAN a jej architektúru CTGAN pre generovanie syntetických údajov.

V závere môžeme konštatovať, že model CTGAN je vhodný pre generovanie tabuľkových údajov. Taktiež sa potvrdilo, že opätovne nepoužíva duplicitné záznamy, ktoré sa vyskytovali v pôvodných údajoch. Ladením hyperparametrov sme sa snažili nájsť najlepšie hodnoty, tak aby sme mali čo najkvalitnejšie syntetické údaje.

Pomocou rozhodovacích stromov sme pomocou metrík, ale aj grafických vizualizácií rozhodovacích stromov porovnávali pôvodné a syntetické údaje, kde sme dosahovali výborné a dobre výsledky pri pôvodných údajoch. Pri syntetických sme si v niektorých prípadoch mohli všimnúť aj nízke alebo nulové výsledky, ktoré mohli vzniknúť z dôvodu, že v pôvodnej množine sa vyskytovalo menej záznamov ako v tej syntetickej, čím vznikali takéto výsledky. Taktiež pri grafickom zobrazení grafov sme si všimli, že niektoré stromy mali podobný a niektorí odlišný tvar vzhľadom na pôvodné a syntetické údaje.

Podobnosť pôvodných a syntetických údajov sme sledovali pomocou jednotlivých vizualizácií. Mohli sme si všimnúť najväčšiu podobnosť pri najväčšom datasete, ktorý obsahoval iba numerické atribúty a taktiež aj pri strednom datasete, najmenej sa podobali údaje pri najmenšom datasete. Vizualizácie potvrdili vysokú podobnosť medzi skutočnými a syntetickými údajmi, a tým aj to, že sú vhodné ochrániť súkromie v medicínskych záznamoch.

Výsledky tejto práce ponúkajú postup, akým je možné bezpečne pracovať s citlivými medicínskymi dátami. V práci sme ukázali, že je možné na základe originálnych dát vytvoriť dostatočne podobné syntetické dáta a využiť syntetické dáta pri vytváraní či testovaní modelu, pričom výsledný model je možné využiť na reálnych medicínskych dátach.

POĎAKOVANIE

Táto práca vznikla vďaka podpore Agentúry na podporu výskumu a vývoja na základe Zmluvy č. APVV-17-0550.

ZDROJE

- [1] Mrinal Kanti Baowaly, Chia-Ching Lin, Chao-Lin Liu, Kuan-Ta Chen . Synthesizing electronic health records using improved generative adversarial networks . [Online] . Journal of the American Medical Informatics Association, March 2019 <https://academic.oup.com/jamia/article/26/3/228/5235390?login=true> .
- [2] Stavroula Bourou, Andreas El Saer, Terpsichori-Helen Velivassaki, Artemis Voulkidis, Theodore Zahariadis. [Online] A Review of Tabular Data Synthesis Using GANs on an IDS Dataset. <https://www.mdpi.com/2078-2489/12/9/375/htm> .
- [3] Edward Choi, Siddharth Biswal, Bradley Malin, Jon Duke, Walter F. Stewart, Jimeng Sun. Generating Multi-label Discrete Patient Records using Generative Adversarial Networks. [Online] Proceedings of the 2nd Machine Learning for Healthcare Conference, 2017 <http://proceedings.mlr.press/v68/choi17a.html> .
- [4] Xu, Lei, S.M. Massachusetts Institute of Technology. Synthesizing tabular data using conditional GAN. [Online] Department of Electrical Engineering and Computer Science, February, 2020 <https://dspace.mit.edu/bitstream/handle/1721.1/128349/1202001437-MIT.pdf?sequence=1&isAllowed=y>
- [5] Noseong Park , Mahmoud Mohammadi , Kshitij Gorde , Sushil Jajodia , Hongkyu Park , Youngmin Kim. Data Synthesis based on Generative Adversarial Networks. [Online] Cornell University, 2017 <https://arxiv.org/abs/1806.03384>

Detekcia kolíznych stavov motorizovaných zostáv

¹Peter GIROVSKÝ, ²Jaroslav MURÍN

^{1,2} Katedra elektrotechniky a mechatroniky, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach

¹peter.girovsky@tuke.sk, ²jaroslav.murin@student.tuke.sk

Abstrakt — Cieľom tejto práce je predstaviť nástroj na detekciu kolíznych stavov motorizovaných zostáv pre špecifické prostredie experimentov realizovaných na synchrotrónoch v DESY, Hamburg. Základom práce je vytvorený program Jastrab, napísaný v jazyku Python, využívajúci fyzikálny simulátor PyBullet a na experiment prepojený cez existujúce rozhranie Kamzik. Funkčnosť programu bola úspešne overená na dvoch testovacích experimentoch v laboratóriách CFEL, v DESY.

Kľúčové slová — Jastrab, detekcia kolízií, detekcia vzdialeností, virtuálna simulácia, 3D model

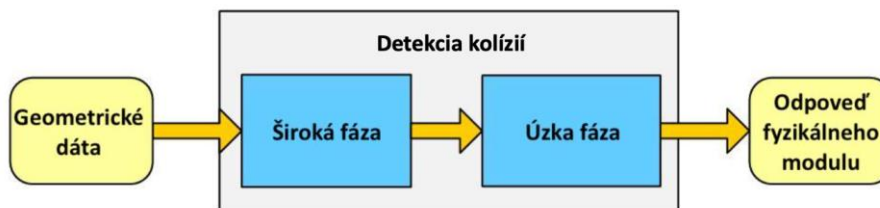
I. ÚVOD

Národný výskumný inštitút DESY (Deutsches Elektronen Synchrotron), v Hamburgu prevádzkuje urýchľovače častíc používané na skúmanie štruktúry hmoty. Jednou z úloh počas týchto experimentov je riadiť zariadenia (väčšinou motory) v motorizovaných zostavách, ktoré sú ich nevyhnutnou súčasťou. V experimentoch sa na odraz alebo lom žiarenia často používajú šošovky, ktorých pozíciu je nutné meniť. Šošovky sú pripevnené na ramenách motorov ktoré sa pohybujú a vzdialenosť medzi nimi je veľmi malá, niekedy dosahuje len mikrometre. Pri takýchto vzdialenostiach je každý pohyb potenciaľne nebezpečný, pri rozbití šošoviek celý experiment skončí zlyhaním. To sa však nemusí týkať len týchto šošoviek, niekedy môže jeden motor naraziť do druhého a zničiť sa tým skúmané biologické vzorky. Z tohto dôvodu bolo potrebné vyvinúť kolízno-detekčný systém ktorý by dokázal tieto zlyhania eliminovať.

II. DETEKCIA KOLÍZIÍ

Hlavnou úlohou modulu na detekciu kolízií je poskytovať informácie o existencii a vlastnostiach interakcií medzi geometrickými tvarmi opisujúcimi simulované objekty. Aby modul na detekciu kolízií fungoval správne, simulačný modul mu musí tieto informácie poskytnúť s aktuálnymi polohami a orientáciami objektov [1]. Tradičný a stále dominantný prístup (v interaktívnej počítačovej grafike) spočíva v testovaní priesečníkov vo vopred určených bodoch v čase, vo virtuálnej scéne. Nazýva sa to diskretná detekcia kolízií (ďalej len DDK). Problém s DDK je v tom, že neexistuje spôsob, ako zabrániť prenikaniu predmetov do seba, a v extrémnom prípade môže viesť k javu známemu ako tunelovanie, keď rýchlo sa pohybujúci objekt prechádza ďalším objektom v po sebe nasledujúcich časových okamihoch bez toho, aby bola zaregistrovaná akákoľvek kolízia. Na zmiernenie tohto problému boli navrhnuté rôzne techniky, ako napríklad kontinuálna detekcia kolízií (ďalej len KDK). Namiesto jednoduchého zaobchádzania so simulovanými objektami, ako s fixovanými v priestore v danom časovom okamihu, KDK analyzuje ich dynamický stav a snaží sa predvídať ich pohyb. Technika KDK je však výpočtovo jednoznačne náročnejšia ako DDK, a preto sa v súčasných implementáciách používa o čosi menej. Jednoduchá implementácia modulu pre detekciu kolízií by testovala každú dvojicu objektov či sa nepretínajú, ktorý v prípade n objektov vyžaduje $O(n^2)$ testov. [4, 5] Aby sa obmedzilo množstvo nepotrebných výpočtov, proces detekcie kolízií sa zvyčajne delí na tieto fázy:

- široká fáza, ktorá používa ohraničovacie obálky, hierarchické delenie priestoru, a iné podobné metódy na vyradenie párov objektov, pri ktorých sa dá preukázať, že objekty ktoré ich tvoria, sa nepretínajú
- úzka fáza, počas ktorej sa vykonávajú výpočtovo náročnejšie skúšky a testy, ale iba medzi pármí, ktoré sa dostali cez širokú fázu. [2, 3]



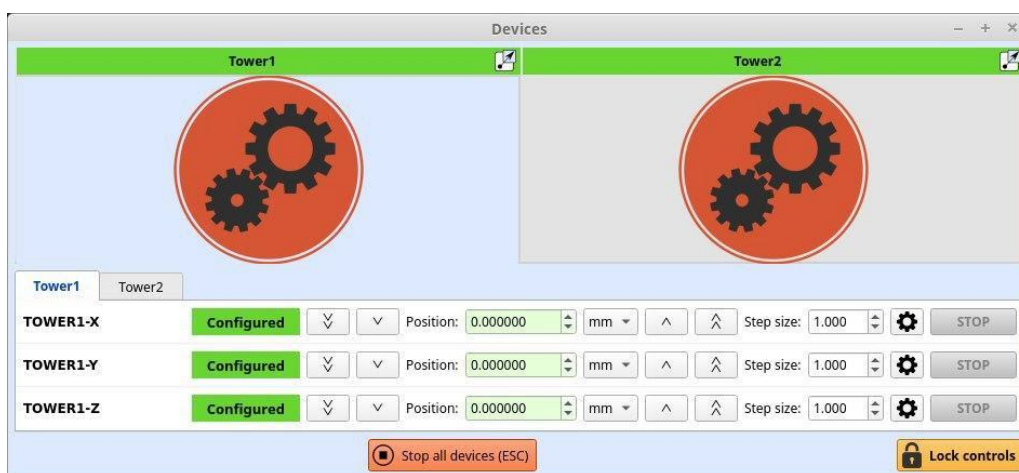
Obr. 1 Priebeh procesu detekcie kolízií

Pre riešenie detekcie kolíznych zostáv boli navrhnuté programy Kamzík a Jastrab, ktoré sú popísané v nasledujúcich častiach článku.

III. KAMZÍK

Počas experimentov sa používa množstvo zariadení, zvyčajne motorov a všetky zariadenia, s ktorými je potrebné počas experimentu komunikovať sú riadené systémom Kamzík. Medzi takéto zariadenia patria motory alebo detektor. Kamzík je prispôsobený na riadenie rôznych typov motorov, akými sú napríklad krokové motory, jednosmerné motory, piezoelektrické motory a pod., pričom ich riadenie je plne asynchrónne, s veľmi veľkou efektívnosťou a presnosťou.

Kamzík sa skladá z dvoch častí - klienta a servera, server je zodpovedný za komunikáciu Kamzík s jednotlivými zariadeniami, ktoré riadi a klient je tá časť programu, z ktorého sa vysielajú riadiace pokyny na server. Grafické okno klienta Kamzík je zobrazené na Obr. 2.

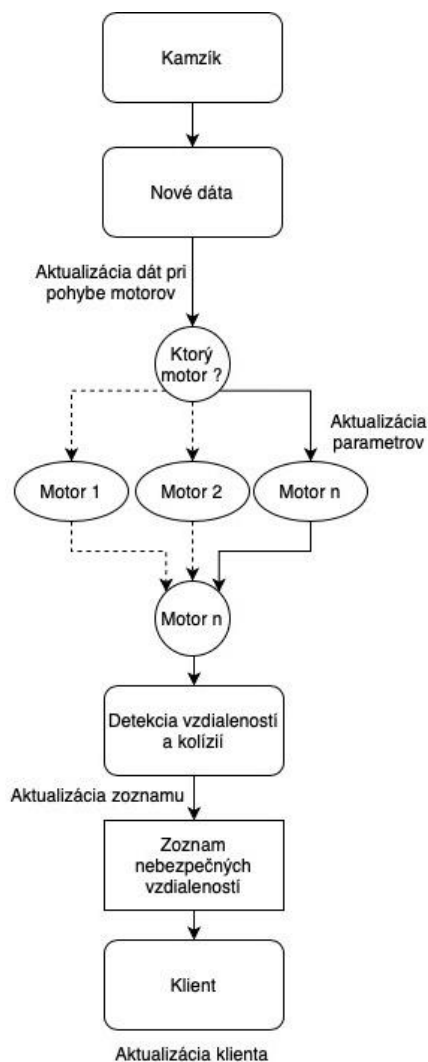


Obr. 2 Grafické okno klienta programu Kamzík

Na obrázku je vidieť, že na testovacom experimente sú použité dve veže, Tower1 a Tower2. Každá z týchto veží obsahuje tri motory, pohybujúce sa v osiach x, y a z. Každý motor sa pohybuje len v jednej z týchto osí. Každý pohyb motorov sa vykonáva po osi buď v kladnom alebo zápornom smere. Veľkosť pohybu (Step size) pri jednom stlačení tlačidla so znakom ^ sa volí ľubovoľne v určených jednotkách. Pohyb motora sa zastaví po stlačení tlačidla STOP. Po stlačení klávesy ESC sa zastaví činnosť všetkých zariadení.

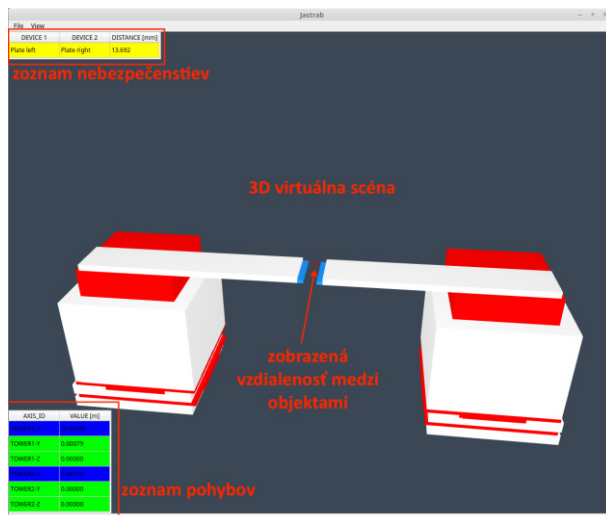
IV. JASTRAB

Program Jastrab je rozdelený na dve časti, server a klient. Program server je hlavná časť programu Jastrab. Má za úlohu simulovať pohyby reálnych motorov vo virtuálnom prostredí. Virtuálnu scénu simuluje z dát ktoré prijíma z Kamzík alebo iného riadiaceho programu a tie potom využíva na detekciu kolízií. Informácie dôležité pre používateľa posielajú ďalej, napríklad na Jastrab klienta, na ktorom sa všetko zobrazuje. Proces spracovania dát programu Jastrab je zobrazený na Obr. 3.



Obr. 3 Vývojový diagram procesu spracovania dát

Klient je časť programu, bez ktorej je Jastrab plne schopný fungovania. Jedným z dôvodov rozdelenia programu Jastrab na server a klienta je, že server môže byť spustený len jeden, ale klientov môže byť naraz spustených ľubovoľný počet a všetky pripojené na jeden server. Klient je určený len na zobrazovanie virtuálnej scény a iných informácií, ktoré môžu byť potrebné pre užívateľa v čase riadenia experimentu, nevie nijak ovplyvňovať server alebo zasahovať do jeho štruktúry dát. Neprebiehajú na ňom žiadne výpočty, vďaka čomu môže byť spustený aj na výkonovo slabšej počítačovej architektúre. Grafické okno programu Jastrab klient je zobrazené na Obr. 4.

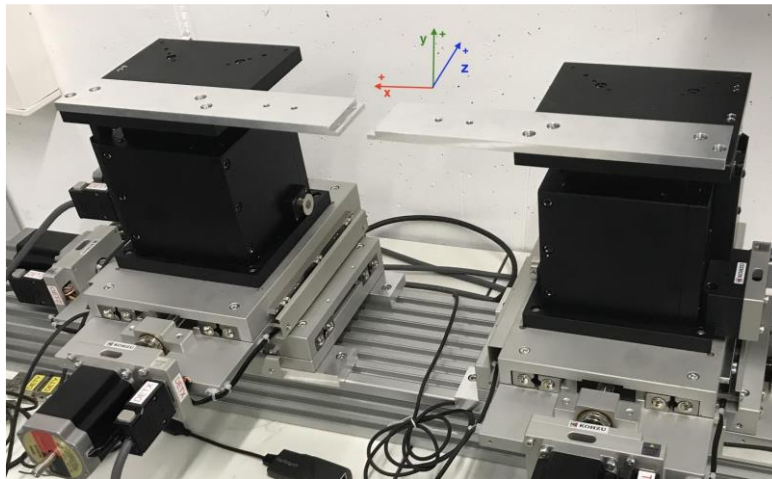


Obr. 4 Grafické okno Jastrab klienta

Grafické okno možno rozdeliť na tri hlavné časti. Vľavo hore sa nachádza zoznam nebezpečenstiev, vľavo dole sa nachádza zoznam hodnôt pohybov jednotlivých pohyblivých objektov(motorov), a grafická časť vykresľujúca všetky objekty virtuálneho prostredia. Po spustení klient pošle na server žiadosť o inicializačné údaje, po ich obdržaní je zosynchronizovaný so serverom. Potom klient čaká na aktualizácie už len meniacich sa údajov, ktoré server vysiela všetkým aktívnym klientom.

V. OVERENIE FUNKČNOSTI PROGRAMU

Overenie funkčnosti programu bolo testované na jednoduchom experimente s dvoma motorizovanými zostavami podľa Obr. 5. Každá zostava sa skladala z troch zariadení (motorov) a jednej kovovej platničky s výrezom.



Obr. 5 Zostava motorov overenia funkčnosti programu

Spodná časť zostavy pozostáva z motora pohybujúceho sa v osi z. Na ňom je postavený motor so stupňom voľnosti v osi y a na vrchu je motor pohybujúci sa v x osi. Testovacou úlohou pri tomto experimente bolo prisunúť kovové platničky čo najbližšie ku sebe bez kolízie.

Test zahŕňal pohyby motormi TOWER1-X a TOWER2-X. Na začiatku bola vzdialenosť medzi platničkami cez 15 mm. Pohli sme TOWER1-X v zápornom smere, vzdialenosť bola 11.84 mm, následne sme pohli motorom TOWER2-X v kladnom smere a tým sme zmenšili vzdialenosť medzi doštičkami na 0.09 mm. Aby sme otestovali funkčnosť programu, v Kamzíku sme nastavili veľkosť kroku pohybu na 1 mm a pri vyslaní žiadosti o pohyb TOWER2-X v kladnom smere, Jastrab vyhodnotil že kovové platničky by sa zrazili, a pohyb zamietol, čím situáciu vyhodnotil správne. Pre otestovanie spoľahlivosti Jastraba bolo vykonaných ešte viacero testov, pričom Jastrab situáciu vždy vyhodnotil správne.

VI. ZÁVER

Cieľom tohto článku bolo predstaviť navrhnutý kontrolný systém, ktorý by mal schopnosť detekovať vzdialenosť medzi motorizovanými zostavami a ochrániť ich pred kolíziami. Takýto kontrolný systém pozostával z programov Kamzík a Jastrab, na základe ich prezentovanej funkčnosti môžeme konštatovať, že navrhnutý systém dokáže zabrániť kolíznym stavom motorizovaných zostáv.

POĎAKOVANIE

Táto práca bola podporovaná Agentúrou na podporu výskumu a vývoja na základe Zmluvy č. APVV-19-0210.

Táto práca vznikla s podporou grantu VEGA 1/0493/19.

LITERATÚRA

- [1] Stępień, Jakub. 2013. Physics-Based Animation of Articulated Rigid Body Systems for Virtual Environments.
- [2] Brian Mirtich. 1998. V-Clip: fast and robust polyhedral collision detection. ACM Trans. Graph. 17, 3 (July 1998), 177–208. DOI:<https://doi.org/10.1145/285857.285860>

- [3] Montanari, M., Petrinic, N., Barbieri, E., 2016. Improving the GJK algorithm for faster and more reliable distance queries between convex objects. *ACM Trans. Graph.*
- [4] Avril, Quentin & Gouranton, Valérie & Arnaldi, Bruno. (2010). A Broad Phase Collision Detection Algorithm Adapted to Multi-cores Architectures. *VRIC 2010 Proceedings.*
- [5] Bajt S, Prasciolu M, Fleckenstein H, et al. X-ray focusing with efficient high-NA multilayer Laue lenses. *Light, Science & Applications.* 2018 ;7:17162. DOI: 10.1038/lsa.2017.162.
- [6] Mark Pilgrim. 2010. *Dive Into Python 3.* CreateSpace, Scotts Valley, CA.
- [7] Erwin Coumans, & Yunfei Bai. (2016–2020). *PyBullet*, a Python module for physics simulation for games, robotics and machine learning. <http://pybullet.org>.

Vplyv rôznych penetrácií domácich FVE na nízkonapäťovú distribučnú sieť

¹Dávid MARTINKO, ²Michal KOLCUN

¹ Katedra elektroenergetiky, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

² Katedra elektroenergetiky, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

¹david.martinko@student.tuke.sk, ²michal.kolcun@tuke.sk

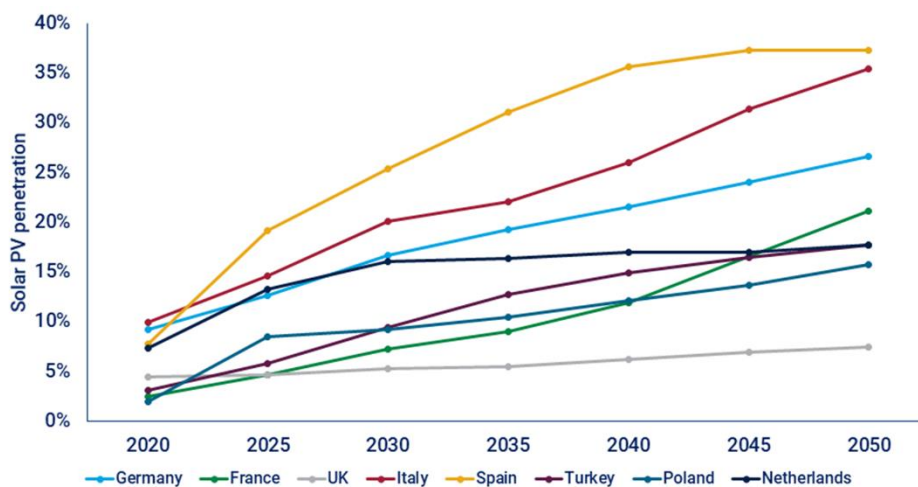
Abstrakt — Tento článok sa venuje matematickému modelovaniu výroby domácich strešných fotovoltaických systémov pri rozdielnych penetráciách v zásobovacej oblasti jednej distribučnej trafostanice na základe historických meteorologických údajov a historických údajov z inteligentných meracích systémov u odberateľov a ich vplyv na plánovaný scenárov EÚ o využití elektriny z obnoviteľných zdrojov do roku 2030.

KLúčové slová — Fotovoltaické systémy, FVE, modelovanie výkonu výroby elektriny, IMS, big data distribučná sústava

I. ÚVOD

Európska únia (ďalej EÚ) a jej všetky členské štáty sa v zmysle Parížskej dohody z roku 2016, ktorej cieľom bolo obmedziť maximálny celosvetový nárast priemernej teploty vplyvom klimatických zmien o 1,5° C a najviac o 2° C, dohodli záväzne znížiť do roku 2030 emisie aspoň o 40% v porovnaní s úrovňami z roku 1990 [1]. Následne v roku 2020 o zníženie emisií o 55% do roku 2030. Cieľom EÚ na rok 2030 je minimálne zdvojnásobenie podielu energie z obnoviteľných zdrojov zo súčasnej úrovne 32 % na približne 65 % alebo viac [2]. Na dosiahnutie tohto cieľa musí EÚ výrazne zvýšiť využívanie obnoviteľných zdrojov energie v sektore energetiky, pričom značná časť z toho bude pochádzať zo solárnych systémov.

Predpokladá sa, že do roku 2050 bude solárna technológia produkovať 20% energií v Európe. To znamená predpokladaný ročné tempo rastu penetrácie fotovoltaických elektrární (ďalej FVE) na úrovni 10%, zobrazené na Obr. 1.



Obr. 1 Predpoklad vývoja zastúpenia FVE na výrobe elektriny vo vybraných krajinách Európy do roku 2050 [3].

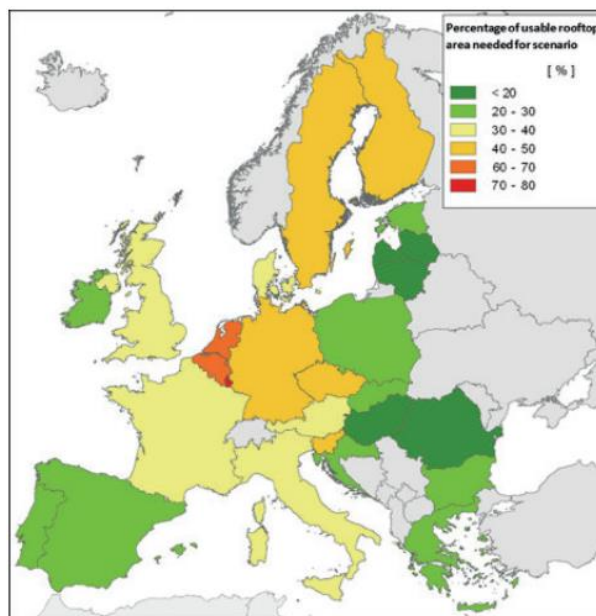
II. ANALÝZA DOMÁCICH FVE NA SLOVENSKU A V EÚ

Percentuálny podiel domácich FVE na Slovensku nie je z oficiálnych štatistík známy. Z bilancie výroby elektriny na Slovensku v roku 2021 vieme povedať, že celkový podiel elektriny vyrobenej z obnoviteľných zdrojov mimo vodných zdrojov je 8% [4]. Na základe úrovne inštalovaného výkonu FVE vieme povedať, že FVE má približne 25% podiel obnoviteľných zdrojov, respektíve 6% celkového inštalovaného výkonu výroby elektriny na Slovensku [5]. Zastúpenie domácich fotovoltaických elektrární nie je známe, vzhľadom k tomu, že v celkovom počte viac ako 2300 FVE na Slovensku sú zarátané aj solárne parky.

Podľa analýzy Spoločného výskumného centra Európskej komisie [6] je potenciál domácich, respektíve FVE inštalovaných na strechách domov a budov na úrovni 680 TWh v celej EÚ, čo predstavuje približne 24,4% celkovej spotreby v roku 2016.

Zatiaľ čo klesajúce investičné náklady a rastúce ceny elektriny by mali podporiť výstavbu nových FVE, potenciál FVE v Európe je stále relatívne nevyužitý. Rozvoj investícií do domácich alebo strešných FVE akcelerujú ak politické rozhodnutia o povinnosti inštalovať FVE na vybrané typy novostavieb komerčných budov od roku 2025 a domácností od roku 2027 v Európskej únii [7].

Pre naplnenie scenárov EÚ využitia elektriny z obnoviteľných zdrojov v roku 2030 a s tým spojeným využitím strešných alebo domácich FVE je potrebné použitie od 20 do viac ako 60% plôch budov, resp. penetrácií FVE na budovách (Obr. 2) [8].



Obr. 2 Potrebné percento dostupnej strešnej plochy potrebnej pre scenár EÚ pre rok 2030 [8]

III. VÝPOČET

A. Model výkonu fotovoltaického systému na základe historických meteorologických dát

Výkon fotovoltaického systému možno opísať ako vzťah (1) medzi poveternostnými podmienkami a menovitým, referenčným (ref) výkonom pri štandardných testovacích podmienkach (STC) [9]

$$P_{PV(t)} = \eta P_{PVref} \frac{G}{G_{ref}} \left(1 + \gamma (T_c - T_{ref}) \right) \quad (1)$$

kde $P_{PV(t)}$ je výkon fotovoltaického systému v časovom intervale, η je účinnosť fotovoltaického systému, respektíve súhrn rôznych strát, G je slnečné žiarenie (W/m^2) prijaté na rovine modulu, γ je teplotná korekcia na maximálny výkon a T_c je teplota panelu. Dolný index „ref“ sa vzťahuje na štandardné testovacie podmienky panelu (STC) $G=1000 W/m^2$, $T=25^\circ C$. P_{PVref} je maximálny výstupný výkon panelu pri referenčných podmienkach.

V reálnych FV systémoch musíme do výpočtov zahrnúť aj straty, ako sú teplotné straty, straty z nesúlady, straty káblov, tiene, sneh, nečistoty a prach, keďže je utopické predpokladať premenu všetkého na modul dopadnutého žiarenia na vyrobenú elektrinu.

Časť slnečného žiarenia sa pomocou solárnych článkov premení na elektrickú energiu. Ostatná

časť sa premieňa na tepelnú energiu. Vzhľadom na to, že solárne články sú polovodičové, sú citlivé na teplotu, ktorú vieme predpovedať viacerými modelmi. Pre výpočet výkonu FVE budeme používať model Mattei 1 [10] pre výpočet teploty panelov

$$T_C = \frac{U_{PV}T_a + I[\tau\alpha - \eta_{STC}(1 - \beta_{STC}T_{STC})]}{U_{PV} + \beta_{STC}\eta_{STC}I} \quad (2)$$

$$U_{PV} = 26.6 + 2.3 v_w$$

kde zohľadníme aj vlastnosti solárnych článkov, ako sú teplotný koeficient pre maximálny výkon β , účinnosť článku η , priepustnosť povrchu τ a absorpčný koeficient článkov α . Vzťah $\tau\alpha$ je definovaný hodnotou 0.81. Koeficient výmeny tepla U_{PV} je funkciou rýchlosti vetra v blízkosti panelov v_w , ktorá sa dá odvodiť z rýchlosti vetra 10m nad úrovňou zeme v_f , dostupnou v historických meteorologických dátach.

$$U_{PV} = 26.6 + 2.3 v_w \quad (3)$$

$$v_w = 0.68 v_f - 0.5$$

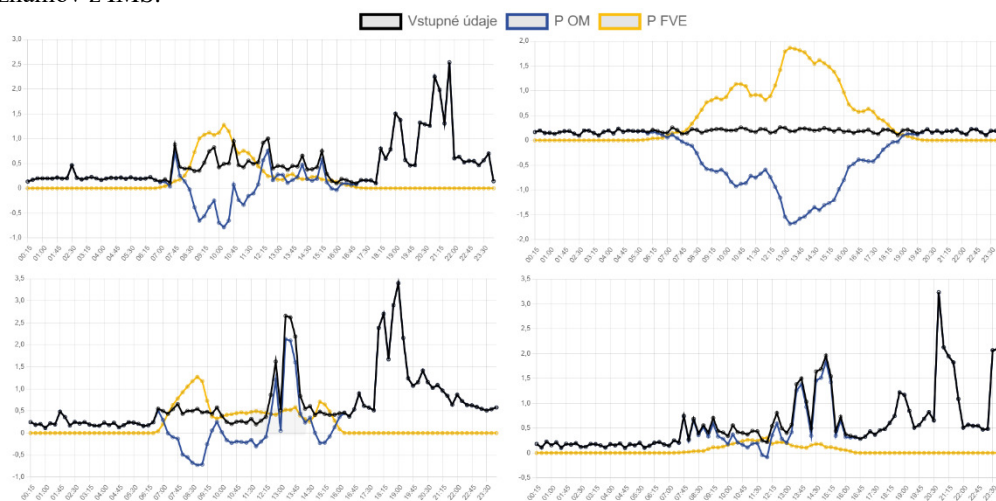
Pri porovnaní rôznych modelov výpočtu výkonu FVE [11] bol model Mattei 1 približne rovnako presný ako model Sklopaki 3, pri ktorom je ale potrebné uvažovať s typom strechy, resp. uložením panelov na nej a keďže výpočet výkonu FVE je vo vybranej lokalite na náhodne vybraných rodinných domov a predpokladá optimálne uloženie panelov na streche, tak uvažujeme len s rýchlosťou resp. intenzitou vetra. Model Mattei je rovnako vhodný pre panely typu polycrystal aj monocrystal. Sklopaki model je vhodnejší pre výpočet s monocrystal technológiou panelov.

Ako zdroj meteorologických dát sme použili dáta od spoločnosti Solcast [12], ktorá poskytuje dáta pre hodnotenie solárnych zdrojov alebo predpovedí pre slnečné žiarenie a výkon FVE na základe globálnej flotily meteorologických satelitov. Dátové súbory obsahujú teplotu, rýchlosť a smer vetra, vlhkosť, oblačnosť a niekoľko typov slnečného žiarenia vypočítaných pre konkrétnu lokalitu na základe GPS súradníc, sklonu a azimutu fotovoltaických panelov v rôznych časových intervaloch.

B. Výpočet zaťaženia distribučnej siete vo vybranej lokalite

Pre porovnanie vplyvu rôznych penetrácií inštalovaných FVE v domácnostiach sme zvolili zásobovaciu oblasť 1 distribučnej trafostanice v obci Malá Ida, ktorá má najvyšší podiel inštalovaných inteligentných meracích systémov (ďalej IMS) medzi trafostanicami prevádzkovateľa distribučnej sústavy na východe Slovenska. Počet OM v uvedenej trafostanici je 97 odberných miest (ďalej OM), z ktorých je 5 bez IMS.

Vstupné dáta z IMS sú s časovým intervalom 15 minút, anonymizované a obsahujú výkon P odoberaný alebo dodávaný do distribučnej siete zo strany zákazníka v časovom intervale. Vstupné dáta pokrývajú roky 2020 a 2021, celkovo sa jedná o big data spravované ako 6,5 mil. záznamov z IMS.



Obr. 3 Príklady denných priebehov spotreby OM pri modelovanom výkone FVE

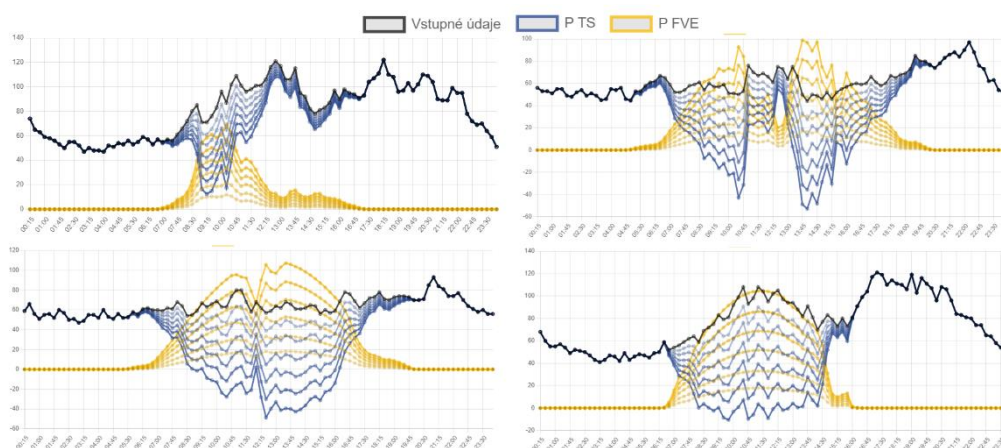
Pre výpočet celkového zaťaženia nízkonapäťovej distribučnej siete pri rôznych penetráciách od 10% do 60% FVE sme náhodne vybrali prislúchajúci počet OM s IMS. Pre tieto vybrané OM sme vypočítali výkon FVE v časovom intervale 15min počas rokov 2020 a 2021 na základe náhodne vybraného systému FVE zo datasetu viac ako 70 vybraných skutočných FVE systémov

z celkového počtu viac ako 500 inštalovaných a evidovaných FVE v distribučnej sústave na východe Slovenska. Súbor skutočných FVE obsahuje typ a počet panelov, parametre panelov, výkon systému, typ a parametre meniča. Príklad denných priebehov spotreby OM pri modelovanom výkone FVE v rôznych ročných obdobiach je znázornený na Obrázku 3.

Celkové zaťaženia distribučnej trafostanice pre jednotlivé 15 minútové časové intervaly boli súčtom zaťaženia OM zo vstupných dát, kde pri vybraných OM s FVE boli nahradené prepočítaným zaťažením OM vrátane modelovaného výkonu FVE. Tieto výpočty pri vzorke dát sme opakovali 12 krát pre zvýšenie náhodnosti výsledkov.

IV. VÝSLEDKY

Na obrázku 4 sú príklady denných priebehov skutočného nameraného výkonu TS označeného ako vstupne údaje, modelovaného výkonu P FVE a modelovaného výkonu TS pri rozdielnych penetráciách FVE označené ako P FVE v rozdielnych ročných obdobiach. Intenzita čiar znázorňuje penetráciu FVE od 10% pri najslabšej po 60% pri najsilnejšej farbe s krokom 10.



Obr. 3 Príklady denných priebehov skutočného nameraného výkonu TS, modelovaného výkonu FVE a modelovaného výkonu TS pri rozdielnych penetráciách FVE

V tabuľke 1 sú priemerné výsledky výpočtov celkovej ročnej spotreby zásobovacieho okruhu vybranej distribučnej trafostanice pri rozdielnych penetráciách domácich FVE a percentuálny rozdiel voči vstupnému stavu bez FVE.

Tabuľka 1
Porovnanie celkovej ročnej spotreby zásobovacieho distribučnej TS

Penetrácia FVE	Spotreba (kWh)		rozdiel voči 0% FVE (%)	
	rok 2020	rok 2021	rok 2020	rok 2021
0% - Vstupné údaje	619 984	646 150	-	-
10%	591 839	618 004	4,54	4,36
20%	455 487	593 025	26,53	8,22
30%	537 920	564 083	13,24	12,70
40%	511 485	537 647	17,50	16,80
50%	484 030	510 192	21,93	21,04
60%	456 687	482 846	26,34	25,28

Uvedený model počíta s FVE bez batériového úložiska so všetkým využitím vyrobenej elektriny v domácnosti alebo prebytkom do distribučnej sústavy.

V. ZÁVER

Cieľom tohto modelovania bolo sledovanie vplyvu inštalácií domových strešných FVE v rozdielnych penetráciách, ktorých rozvoj je očakávaný aj plánovaný v zmysle scenárov EÚ o využití elektriny z obnoviteľných zdrojov do roku 2030. Pri nadpolovičnej penetrácii FVE môže dôjsť v celkovej ročnej spotrebe elektriny až k štvrtinovému poklesu voči distribučnej sieti bez domácich FVE. Budúcim smerom je využívanie FVE v kombinácii s batériovým úložiskom, ktoré má prínos pre odberateľov vo využití vyrobenej elektriny v čase mimo slnečného svitu ale súčasne znižujú riziko negatívnej spotreby v celej zásobovacej oblasti distribučnej trafostanice v letných mesiacoch, respektíve v obdobiach vysokého slnečného žiarenia a nižšieho celkového

zaťaženia trafostanice. V tabuľke 2 sú priemerné počty dní so celkovou zápornou spotrebou pre zásobovacieho okruhu trafostanice aspoň v jednom 15 minútovom časovom intervale z uvedeného dňa.

Tabuľka 2
Priemerný počet dní so zápornou spotrebou zásobovacieho okruhu trafostanice
aspoň v jednom 15 minútovom časovom intervale

Penetrácia FVE	Počet dní	
	rok 2020	rok 2021
10%	0	0
20%	0	1
30%	27	17
40%	31	30
50%	31	31
60%	31	31

POĎAKOVANIE

Tento príspevok bol podporený Agentúrou na podporu výskumu a vývoja na základe zmluvy č. APVV-19-0576, ako aj Vedeckou grantovou agentúrou Ministerstva školstva, vedy, výskumu a športu SR a SAV, podporený grantom VEGA 1/0757/21.

LITERATÚRA

- [1] Európsky parlament a rada Európskej únie, "Smernica európskeho parlamentu a rady (eú) 2018/2001 z 11. decembra 2018 o podpore využívania energie z obnoviteľných zdrojov" Úradný vestník Európskej únie, 2018.
- [2] Európska komisia, "Oznámenie komisie európskemu parlamentu, rade, európskemu hospodárskemu a sociálnemu výboru a výboru", <https://eur-lex.europa.eu/legal-content/SK/TXT/PDF/?uri=CELEX:52020DC0562&from=EN>, 2020.
- [3] Wood Mackenzie: Sunny side up: our European solar PV market outlook 2021. [online]. Wood Mackenzie. [citované 13.5.2022]. Dostupné na internete, <https://www.woodmac.com/news/opinion/sunny-side-up-our-european-solar-pv-market-outlook-2021/>
- [4] Energoklub: Na Slovensku sa vlani vyrobilo viac elektriny. [online]. Energoklub.sk. [citované 13.5.2022]. Dostupné na internete, <https://energoklub.sk/sk/clanky/na-slovensku-sa-vlani-vyrobilo-viac-elektliny-tahunom-bol-zemny-plyn/>
- [5] Výroba elektrickej energie na Slovensku a jej zdroje. [online]. CenyEnergii.sk. [citované 13.5.2022]. Dostupné na internete, <https://cenyenergii.sk/2021/06/01/vyroba-elekttrickej-energie-na-slovensku-a-jej-zdroje/>
- [6] K. Bódis, I. Kougias, A. Jäger-Waldau, N. Taylor, S. Szabó, "A high-resolution geospatial assessment of the rooftop solar photovoltaic potential in the European Union" *Renewable and Sustainable Energy Reviews*, volume 114, 2019.
- [7] PV magazine: EU wants rooftop PV mandate for public, commercial buildings by 2027, residential by 2029. [online]. Pv-magazine.com. [citované 20.5.2022]. Dostupné na internete, <https://www.pv-magazine.com/2022/05/18/eu-wants-rooftop-pv-mandate-for-public-commercial-buildings-by-2025-residential-by-2029/>
- [8] T. Huld, K. Bodis, I. P. Pascua, E. Dunlop, N. Taylor, A. Jäger-Waldau, "The Rooftop Potential for PV Systems in the European Union to deliver the Paris Agreement" *European Energy Innovation*, 2018.
- [9] J.I. Rosell, M. Ibanez, "Modelling power output in photovoltaic modules for outdoor operating conditions", *Energy Conversion and Management* 47, 2006.
- [10] M. Mattei, G. Notton, G. Cristofari, M. Muselli, P. Poggi, "Calculation of the polycrystalline PV module temperature using a simple method of energy balance.", *Renew Energy* 2006.
- [11] D. Martinko, "Modelling photovoltaic system power output based on historical Meteorological data", Technical University of Košice, 22nd Scientific Conference of Young Researchers, 2022.
- [12] Solcast, "Solcast API Toolkit" [Online], Dostupné na internete, <https://solcast.com/solar-data-api/api-toolkit/>

Skúmanie modulačných formátov pre Ultra-DWDM prenosové systémy

¹Luboš OVSEŇÍK, ²Jakub ORAVEC, ³Norbert ZDRAVECKÝ,
⁴Samuel ANDREJČÍK, ⁵Maroš LAPČÁK

^{1,2,3,4,5} Katedra elektroniky a multimediálnych telekomunikácií, Fakulta elektrotechniky
a informatiky, Technická univerzita v Košiciach, Slovenská republika

¹lubos.ovsenik@tuke.sk, ²jakub.oravec@tuke.sk, ³norbert.zdravecky@tuke.sk,
⁴samuel.andrejcik@tuke.sk, ⁵maros.lapcak@tuke.sk

Abstrakt — Predmetom tohto príspevku je návrh a analýza 32-kanálového Ultra-DWDM systému s adaptáciou rôznych vysokospektrálnych modulácií, ako nástrojom na regulovanie nežiaduceho vplyvu lineárnych a nelineárnych javov, ktoré sú hlavným faktorom, ktorý limituje kapacitu a prenosovú vzdialenosť optického komunikačného systému.

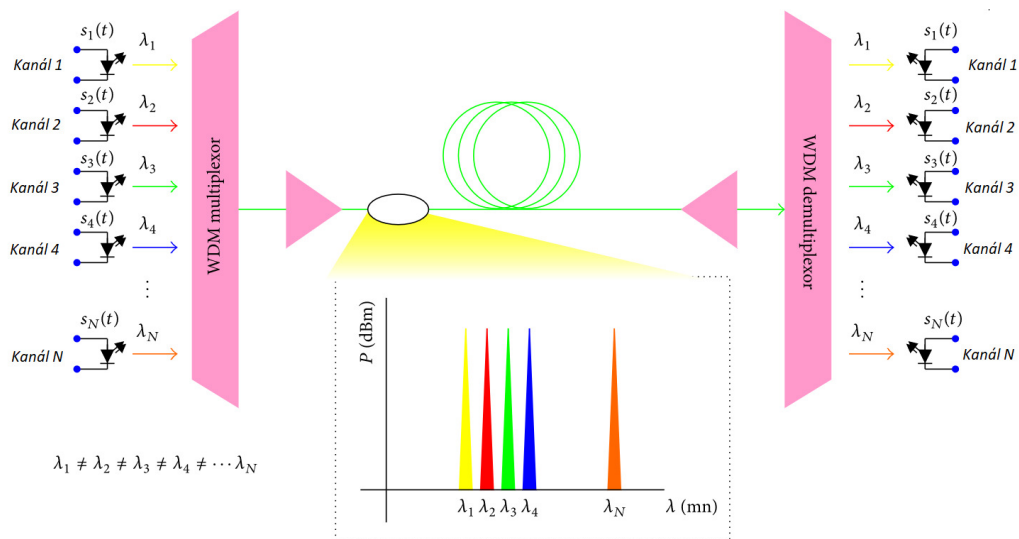
Kľúčové slová — DWDM, optické modulácie, optické pasívne siete, PON, WDM

I. ÚVOD

Informačno-komunikačné technológie dnes dominujú v životoch ľudí a neutíchajúce požiadavky na dostupnosť online služieb a technológií stavajú výskumníkov a konštruktérov do nových výziev. Vďaka požiadavke mať obľúbené online služby dostupné kedykoľvek a kdekoľvek sa prenosové komunikačné systémy v posledných dvadsiatich rokoch vo veľkej miere pretransformovali. Dnes, z pohľadu spoľahlivosti a rýchlosti prenosu, sú optické vláknové komunikačné systémy vedúcou technológiou na poli prenosových systémov. Optické vláknové komunikačné systémy dnes tvoria chrbticové siete a v podstate na nich závisí chod všetkých online služieb. To, čo robí optické vlákno takým neprekonateľným prenosovým médium v porovnaní s metalickými či bezdrôtovými médiami, je šírka pásma, ktorá je prakticky neobmedzená. Keďže sa dáta prenášajú vo forme svetelnej vlny, nie sú signály prenášané optickým vláknom ovplyvňované elektro-magnetickým rušením. Vlnové multiplexovanie WDM (Wavelength Division Multiplexing) je technológia, ktorá umožnila prenos obrovského množstva dát vysokou prenosovou rýchlosťou jediným optickým vláknom. Exponenciálne rastúci dopyt po prenosovej kapacite vo vysokorýchlostných prenosových systémoch na dlhé vzdialenosti prinútil optickú vláknovú komunikačnú technológiu, aby si ako životaschopné riešenie osvojila systém multiplexovania s hustým vlnovým delením DWDM (Dense Wavelength Division Multiplexing). Pre maximalizáciu výkonu takýchto optických sietí sa kritickou oblasťou výskumu stali vhodné stratégie návrhu komunikačného systému na optimalizáciu parametrov súvisiacich s prenosom dát v rámci vlastností kanála. Kapacita optického kanála je zvyčajne ovplyvnená tvarom signálu, disperznými a nelineárnymi charakteristikami optického vlákna.

II. OPTICKÉ VLÁKNOVÉ KOMUNIKAČNÉ SYSTÉMY S HUSTÝM VLNOVÝM DELENÍM (DWDM)

V dvadsiatom prvom storočí sú WDM systémy neodmysliteľnou súčasťou chrbticových optických prenosových sietí, keďže pomocou nich možno prenášať stovky optických signálov s rôznymi optickými vlnovými dĺžkami po jednom optickom vlákne (OV). Celý systém obsahuje n optických zdrojov a n optických detektorov, pričom na každú použitú vlnovú dĺžku vo vysielacej jednotke je namodulovaný optický signál (Obr. 1) [1, 2, 3]. Vysielače sú tvorené zdrojmi optického žiarenia, pričom najčastejšie sa používajú DFB (Distributed Feedback) lasery, pretože ich spektrálna šírka sa pohybuje v jednotkách MHz. DFB využíva rezonátor, ktorý je tvorený Braggovou mriežkou s periodickou zmenou indexu lomu, vďaka čomu generuje úzko spektrálne žiarenie, ktorého šírka neprekračuje šírku jedného kanála DWDM. Zlučovanie všetkých vlnových dĺžok do jedného OV sa deje v bloku multiplexora. Opačná funkcia, teda vydeľovanie jednotlivých vlnových dĺžok, prebieha v bloku demultiplexora [1, 2, 3].



Obr. 1 Schéma vlnového multiplexu.

DWDM, tzv. „hustý“ vlnový multiplex, využíva minimálne odstupy medzi jednotlivými vysielacími kanálmi (12,5 GHz, 25 GHz, 50 GHz, 100 GHz), vďaka čomu dokáže implementovať do jedného OV desiatky až stovky optických vlnových dĺžok. Pri praktických realizáciách je potrebné počítať s tým, že dosah jednotlivých kanálov bude rôzny a zároveň brať do úvahy aj najhoršie možné prenosové vlastnosti daného kanála v multiplexe (popríklad sa snažiť nevyužívať najhoršie kanály) [2, 4]. Hlavnou podmienkou komerčného nasadenia DWDM do chrbticových optických sietí bolo zvládnutie výroby kľúčových komponentov ako EDFA, AWG multiplexných a demultiplexných jednotiek a úzkopásmových laserových vysieláčov. Vďaka pokroku štúdiá v oblasti vývoja a výroby a tiež zvládnutiu ekonomických aspektov pri výrobe bolo možné ich komerčné nasadenie. Na územie Slovenskej republiky sa DWDM dostal začiatkom roka 2008. Pochádzal od spoločnosti CISCO a zaviedla ho firma Slovanet. Normalizovaná referenčná frekvencia, od ktorej sa odvíjajú frekvencie ostatných kanálov, je frekvencia 193,1 THz. Tieto systémy môžu pracovať v pásmach S, C a L. Odporúčanie ITU-T G.694.1 – „Spectral grids for WDM applications: DWDM frequency grid“ definuje jednotlivé prenosové kanály v oblasti vlnových dĺžok v rozsahu od 1 490 nm (200,95 THz) do 1 620 nm (186 THz). Toto odporúčanie počíta s odstupom jednotlivých kanálov v rozsahu 100 GHz, a to so začiatkom na 186 THz (odstup cca 0,8 nm) alebo s dvojnásobným počtom kanálov s odstupom 50 GHz (odstup cca 0,4 nm).

Šírka pásma optického signálu prenášaného v jednom kanáli je závislá od šírky pásma pôvodného modulačného signálu a môže byť ďalej zhoršená nevhodnou modulačnou technikou. Medzi popredných výrobcov DWDM môžeme zaradiť Lucent Technologies, Alcatel, Nortel a NEC.

III. MODULÁCIE OPTICKÉHO ŽIARENIA POUŽÍVANÉ V DWDM SYSTÉMOCH

Vplyvy nelineárnych javov, ako sú XPM (Cross Phase Modulation) a FWM (Four Wave Mixing), môžu byť v konvenčných optických komunikačných systémoch riadené prostredníctvom reziduálnej lokálnej disperzie optického vlákna alebo nastavením rozstupov kanálov na dostatočne veľkú hodnotu. Nahromadená disperzia, ako lineárny jav, môže byť účinne kompenzovaná buď periodicky pozdĺž optického vlákna, alebo na strane prijímača. Obidva spôsoby však zvyšujú zložitosť systému. Ďalšou alternatívou je zväčšenie medzikanálových rozstupov, čo však zhoršuje využitie šírky pásma a vedie k zníženiu účinnosti DWDM systému. Hlavným účelom rozmiestnenia kanálov dostatočne ďaleko od seba a použitia optických vlákien s primeranou disperziou je potlačenie degradačných mechanizmov rezultujúcich do vzniku nových spektrálnych zložiek alebo prekryvaniu sa spektrálnych zložiek prenášaných kanálov. Optická disperzia môže byť účinne kompenzovaná, avšak práve nelineárne javy spôsobujú v DWDM systémoch operujúcich s vysokou prenosovou rýchlosťou a na veľkú vzdialenosť značné problémy. Zo štúdiá nelineárnych javov bolo zistené, že generovanie nelineárnych javov SPM (Self Phase Modulation), XPM a FWM môže byť minimalizované, ak má prenášaný signál isté špeciálne charakteristiky. To vedie k zaujímavej možnosti využitia pokročilých optických modulačných formátov vo viackanálových DWDM systémoch za účelom zmiernenia vplyvu nelineárnych javov [5, 6]. V Tabuľke 1 je porovnanie vlastností vybraných pokročilých

modulačných formátov využívaných v DWDM systémoch.

Tabuľka 1
Hlavné vlastnosti modulačných formátov používaných v 100 Gb/s DWDM systémoch

Modulačný formát	OOK	OOK-VSB	DQPSK	RZ-DQPSK-3ASK	PM-DQPSK	OP-FDM-RZ-DQPSK	PM-QPSK	PM-OFDM-QPSK
Koherentnosť/ nekoherentnosť	nek.	nek.	nek.	nek.	nek.	nek.	koh.	koh.
Bit/symbol	1	1	2	2,5	2x2	2x2	2x2	2x2x2
Symbolová rýchlosť	112	112	56	44	28	28	28	14
Spektrálna účinnosť (bit/s/Hz)	0,5	1	1	2	2	1	2	2

Za účelom minimalizovania lineárnych a nelineárnych javov vo vysoko kapacitných DWDM systémoch je potrebné nájsť optimálny modulačný formát. Modulačný formát s úzkou šírkou pásma môže zvýšiť spektrálnu efektívnosť a odolnosť voči chromatickej disperzii. Naopak, modulačný formát s konštantným optickým výkonom môže byť menej náchylný na vplyv SPM a XPM. Modulačný formát s viacerými úrovňami (multilevel) dokáže obsiahnuť väčšie množstvo informácií ako binárny signál a vďaka dlhšiemu trvaniu symbolu znížiť degradáciu vplyvom chromatickej a polarizačnej disperzie. Navyiac, v diaľkových prenosoch je dôležitým faktorom ovplyvňujúcim vznik nelineárnych javov aj zosilnenie optického signálu často realizované erbiom dopovaným optickým zosilňovačom (EDFA, Erbium Doped Fiber Amplifier), ktorý taktiež vnáša do systému šum a za určitých okolností (dĺžka erbiom dopovaného vlákna, miera zosilnenia) môže v optickom vlákne vybudovať nelineárne javy SPM a XPM. Tento fakt teda musí byť pri návrhu DWDM systému s pokročilou a efektívnou optickou moduláciou zohľadnený [6]. Ďalším typom v praxi používaných dopovaných vláknových zosilňovačov je yterbiom dopovaný optický vláknový zosilňovač (YDFA, Ytterbium Doped Fiber Amplifier).

IV. NÁVRH MODULAČNÝCH FORMÁTOV PRE ULTRA-DWDM SYSTÉM

Jedným z dôležitých problémov pri návrhu systému je porozumieť kódovanej impulzovej odozve v kanáli, ktorá má určité špecifické vzory. Optická spektrálna šírka pásma signálu, tolerancia voči chromatickej disperzii (CD), odolnosť proti nelineárnemu presluchu, citlivosť na nahromadený šum a ďalšie opatrenia týkajúce sa výkonu systému sú v skutočnosti priamo spojené s formátom optickej modulácie. Aby sa minimalizovali lineárne aj nelineárne javy prenosového vlákna, je potrebný optimálny modulačný formát. Modulačný formát s úzkym optickým spektrom môže zvýšiť spektrálnu účinnosť a tolerovať väčšie skreslenie vplyvom CD. Modulačný formát s konštantným optickým výkonom môže byť menej citlivý na SPM a XPM. Naopak modulačný formát s viacerými úrovňami signálu bude prenášať viac informácií ako binárne signály a jeho dlhšie trvanie symbolov zníži skreslenie vyvolané CD a PMD (Polarization Mode Dispersion).

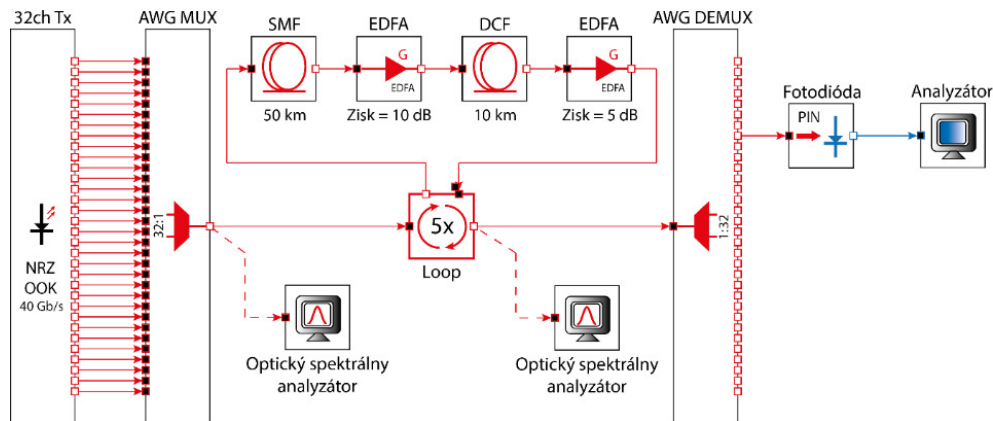
A. Intenzitné modulačné formáty pre DWDM systémy

Pre DWDM systémy operujúce s prenosovou rýchlosťou 40 Gb/s na kanál sme zvolili nasledovné modulačné formáty: NRZ-OOK (Non Return to Zero On-Off Keying), RZ-OOK (Return to Zero On-Off Keying), CSRZ-OOK, Duobinary a modifikovaný Duobinary. CSRZ-OOK (Carrier Suppressed Return to Zero On-Off Keying) je pseudo-viacúrovňový modulačný formát a vyznačuje sa obrátením znamienka optického poľa pri každom bitovom prechode. Znižuje vplyv nelineárnych javov v kanáli a zlepšuje spektrálnu účinnosť v systémoch s vysokou bitovou rýchlosťou. Optické duobinárne a modifikované duobinárne modulačné formáty sa tiež ukázali ako rozumné voľby, pretože majú vysokú toleranciu voči CD, lepšiu nelineárnu toleranciu a môžu prejsť úzkopásmovým optickým filtrovaním, sú teda vhodné pre medzikanálové odstupy menšie ako 100 GHz, aby sa zlepšila spektrálna účinnosť. Ďalej je ľahké ich implementovať, pretože zmeny vo vysielači sú relatívne malé a prijímač ostáva nezmenený – IM/DD (Intensity Modulation–Direct Detection). Celkové nastavenie systému s intenzitnými modulačnými formátmi (NRZ/RZ, CSRZ, Duobinary) je jednoduchšie v porovnaní s inými konkurenčnými formátmi, ako sú DPSK (Differential Phase Shift Keying) a DQPSK (Differential Quadrature Phase Shift Keying), ktoré vyžadujú úpravy na vysielači a prijímači, čo ďalej zvyšuje systémové náklady a zložitosť. Na druhej strane tieto formáty fázovej modulácie umožňujú prenosy na väčších prenosových rýchlostiach a taktiež dosahujú väčšiu spektrálnu účinnosť. Znižujú tiež účinky SPM a XPM, čo vedie k zníženiu skreslenia signálu, ktoré sa hromadí v optickom signáli prenášanom cez optické vlákno, a preukázalo sa, že je silným kandidátom pre systémy Ultra-

DWDM. Systémy DWDM s prenosovou rýchlosťou 112 Gb/s sme analyzovali s modulačnými formátmi DPSK a DQPSK s rôznymi modifikáciami na dosiahnutie požadovaných vlastností za účelom zvýšenia spektrálnej účinnosti a potlačenia nelineárnych javov.

B. Simulačný model pre 1,28 Tb/s DWDM systém

Navrhovaný 32-kanálový DWDM systém pozostáva z vysielačej časti, vláknovej prenosovej časti a optických prijímačov, ako je to znázornené na Obr. 2, pričom stredná frekvencia prvého kanála je 193,0 THz.



Obr. 2 DCF kompenzácia chromatickej disperzie.

Vysielač DWDM pozostáva z generátora pseudonáhodnej bitovej postupnosti (PRBS), CW (Continuous Wave) laserov, optických a elektrických modulátorov a optického multiplexora. Generátor PRBS generuje bitové sekvencie rýchlosťou 40 Gb/s s 2^7-1 bitmi. Emisné frekvencie CW lasera sú rovnomerne rozmiestnené a nachádzajú sa v rozmedzí 193,0 – 196,1 THz s frekvenčným odstupom 100 GHz medzi susednými kanálmi [7]. Hlavný typ modulátora je MZM (Mach-Zehnder Modulator). Extinkčný pomer MZM je nastavený na 30 dB. Vysielačový výkon je 6 dBm. Ku každému výstupnému portu každého CW lasera bol pripojený vysielač NRZ/RZ/CSRZ/Duobinary. Optické signály z 32 takýchto dátových modulátorov sa privádzajú na 32 vstupných portov optického multiplexora AWG (Arrayed Waveguide Grating). Aby sa zabezpečilo oddelenie medzi kanálmi vo frekvenčnej oblasti (lineárne potlačenie krížového presluchu), pred multiplexovaním sa každý kanál opticky filtruje optickým filtrom s úzkym prenosom. Tu sme uvažovali použitie Gaussového filtra druhého rádu so šírkou pásma 100 GHz. Kanálové rozostupy a prevádzkové vlnové dĺžky sú definované v štandardoch ITU-T. Kombinovaný optický signál sa privádza do SMF (Single Mode Fiber) (Tabuľka 2).

Tabuľka 2
Parametre navrhnutého modelu

Parameter	HODNOTA
Bitová rýchlosť	40 Gb/s
Počet vzoriek na 1 bit	64
Počiatočná frekvencia	193,0 THz
Počet kanálov	32
Medzikanálové rozostupy	100 GHz
Vyžarovací výkon	6 dBm
Útlm SMF	0,2 dB/km
Dĺžka 1 segmentu SMF	50 km
Disperzia - β_2	20 ps ² /km
Efektívna plocha SMF A_e	80 μm^2
Nelineárny index lomu n_2	2,6e-20 m ² /W
Simulovaná dĺžka	500 km

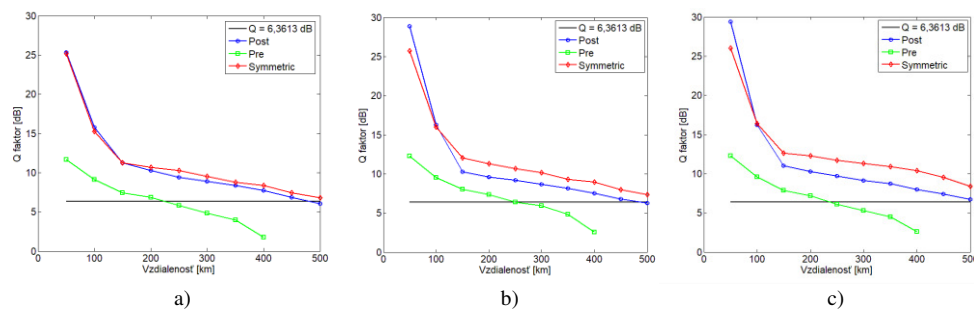
Navrhovaný systém DWDM bol navrhnutý pre kompenzáciu disperzie pre-compensation, post-compensation a symmetrical-compensation. V schéme pre-compensation sa na kompenzáciu disperzie a nelinearít používa vlákno DCF (Dispersion Compensating Fiber) s

dĺžkou 10 km pred vláknom SMF s dĺžkou 50 km. V linke boli použité aj dve in-line EDFA zosilňovače so ziskom 5 dB a 10 dB. Pri schéme post-compensation sa po kompenzácií s nahromadenou disperziou používa vlákno DCF s dĺžkou 10 km po vlákne SMF s dĺžkou 50 km. V schéme symmetrical-compensation sa používa vlákno DCF s dĺžkou 10 km v strede vlákna SMF s dĺžkou 50 km. Tu sa použili tri EDFA so ziskom 5 dB, 10 dB a 5 dB.

V prijímači je signál de-multiplexovaný a detegovaný detektorom PIN. Použitý optický AWG demultiplexor má 32 výstupných portov. Optický signál z každého portu sa potom vedie fotodiódou PIN, ktorej referenčná frekvencia je v rozsahu od 193,0 do 196,1 THz, citlivosť [A/W] = 1 a hodnota prúdu za tmy = 0,1 nA. Potom bol použitý regenerátor 3R na regeneráciu elektrického signálu pripojeného priamo k analyzátoru BER, ktorý sa používa ako vizualizátor na generovanie grafov a výsledkov, ktoré voláme diagramy oka. Z týchto diagramov boli vypočítané hodnoty BER a Q-faktor. Je nutné dodať, že pri analýze systému sa výsledky prvého kanála považovali za výsledky, ktoré zodpovedajú najhoršiemu prípadu, a teda tieto výsledky boli ďalej analyzované.

C. Vyhodnotenie simulácií pre 1,28 Tb/s DWDM systém

Vplyv CD kompenzácie od Q-faktoru navrhnutého systému s NRZ-OOK moduláciou je zobrazený na Obr. 3a. Prah pre modulácie typu OOK pre BER a Q-faktor je: BER = $1e-10$, Q = 6,3613 dB. Na základe toho sme určili, že maximálna dosiahnuteľná vzdialenosť prenosu 1,28 Tb/s DWDM systému s NRZ-OOK s kompenzáciou CD pre-compensation je 200 km s hodnotami BER = $3,75E-12$, Q = 6,8478 dB. Pri metódach post-compensation a symmetrical-compensation bola maximálna dosiahnutá prenosová vzdialenosť 450 km pre post-compensation (BER = $4,12E-12$, Q = 6,8343 dB) a 500 km pre symmetrical-compensation (BER = $4,17E-12$, Q = 6,8326 dB).

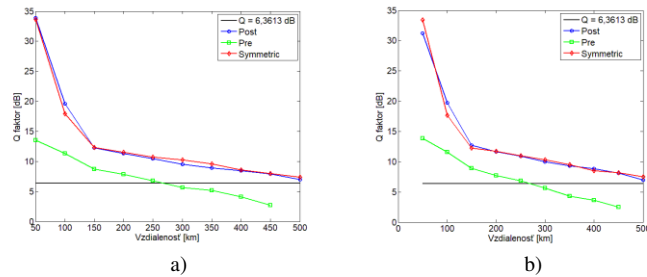


Obr. 3 Q-faktor vs prenosová vzdialenosť pre rôzne typy intenzitných modulácií: (a) NRZ, (b) RZ a (c) CSRZ.

Porovnanie metód kompenzácie CD pre RZ-OOK v 1,28 Tb/s DWDM systéme je na Obr. 3b. Rovnako ako v prípade NRZ-OOK, metóda pre-compensation vykazuje najnižšie hodnoty Q-faktora, a teda je aj v tomto prípade táto metóda kompenzácie najmenej účinná (maximálna prenosová vzdialenosť 250 km, BER = $8,43E-11$, Q = 6,3875 dB). Najúčinnější metóda kompenzácie je symmetrical-compensation s dosiahnutými hodnotami na vzdialenosť 500 km, BER = $7,63E-14$, Q = 7,3849 dB. Pri metóde post-compensation sme dosiahli maximálnu prenosovú vzdialenosť 500 km a hodnoty BER = $2,16E-10$ a Q = 6,2421 dB.

Porovnanie metód kompenzácie CD pre CSRZ-OOK v 1,28 Tb/s DWDM systéme je na Obr. 3c. Rovnako ako v prípade NRZ-OOK a RZ-OOK, môžeme pozorovať, že charakter výsledkov je veľmi podobný. Metóda pre-compensation vykazuje najnižšie hodnoty Q-faktora, a teda je aj v tomto prípade táto metóda kompenzácie najmenej účinná (maximálna prenosová vzdialenosť 200 km, BER = $3,33E-13$, Q = 6,1110 dB). Najúčinnější metóda kompenzácie je symmetrical-compensation s dosiahnutými hodnotami na vzdialenosť 500 km, BER = $2,11E-17$, Q = 8,4063 dB. Pri metóde post-compensation sme dosiahli maximálnu prenosovú vzdialenosť 500 km a hodnoty BER = $9,82E-12$ a Q = 6,7087 dB.

Prenosový výkon na rôznych vzdialenostiach, zobrazený na Obr. 4a pre DRZ (Duobinary RZ) a Obr. 4b pre MDRZ (Modified Duobinary RZ), odhaľuje, že v prípade dokonalej kompenzácie duobinárny formát vykazuje najlepší výkon v schéme pre-compensation. Pri schéme symmetrical-compensation tento DWDM systém dosahuje porovnateľne podobné výsledky ako DWDM systém so schémou post-compensation.



Obr. 4 Q-faktor vs prenosová vzdialenosť pre intenzitné modulácie (a) Duobinary, (b) Modified Duobinary.

V prípade kompenzačnej schémy pre-compensation je maximálna prenosová vzdialenosť 250 km pre moduláciu DRZ: BER = $7,41E-12$, Q = 6,7497 dB a pre MDRZ: BER = $3,22E-12$, Q = 6,8696 dB. Oveľa účinnejšími sú metódy kompenzácie post-compensation a symmetrical-compensation. Výsledné hodnoty BER a Q-faktor pre schému post-compensation na vzdialenosť 500 km sú: BER = $1,77E-12$, Q = 6,9544 dB pre DRZ a BER = $1,21E-12$, Q = 7,0081 dB pre MDRZ. Schéma symmetrical-compensation je aj v tomto prípade na vzdialenosť 500 km najúčinnejšia s hodnotami BER = $7,77E-14$, Q = 7,3825 dB pre DRZ a BER = $3,61E-14$, Q = 7,4838 dB pre MDRZ. Modifikovaná duobinárna modulácia tak mierne vylepšuje výkonnosť navrhnutého DWDM systému vo všetkých troch kompenzačných schémach chromatickej disperzie.

V. ZÁVER

Na základe dosiahnutých výsledkov možno konštatovať, že zmenou tvaru vstupného signálu je možné vo veľkej miere ovplyvniť vznik a vplyv GVD (Group Velocity Dispersion) a nelineárnych javov v DWDM systémoch. Výskum v tejto oblasti je nesmierne dôležitý a umožňuje dosahovať stále vyššie prenosové kapacity a prenosové vzdialenosti DWDM systémov. Existujúce optické prenosové systémy využívajúce OOK (On-Off Keying) modulácie môžu byť pomerne jednoducho modifikované použitím DRZ alebo MDRZ, bez potreby zmeny prijímacích modulov. Takýmto spôsobom je možné prekonať väčšie vzdialenosti vďaka vyššej odolnosti voči XPM a FWM. Rovnakým spôsobom možno dosiahnuť lepšie prenosové charakteristiky aplikovaním RZ-DPSK modulácie, ktorá poskytuje značné zlepšenie odolnosti voči GVD, XPM a FWM v systémoch DWDM s EDFA.

POĎAKOVANIE

Táto práca bola podporená výskumným projektom FEI-2022-84 "Techniky spracovania dát vo vysokorýchlostných prenosových systémoch".

LITERATÚRA

- [1] V. Sharma, et al., "Passive Optical Network: A New Approach In Optical Network", In: *2020 International Conference on Advances in Computing, Communication & Materials (ICACCM)*, IEEE 2020, pp. 295-300. doi:10.1109/ICACCM50413.2020.9213059
- [2] ITU-T G.984.1. Gigabit-capable passive optical networks (GPON): General characteristics. ITU-T. March 2008.
- [3] D&O Committee. 2017. FTTH Handbook. Edition 8. FTTH Council Europe. 204 pp. Revision 2018-03-13.
- [4] ITU-T G.987.3. 10-Gigabit-capable passive optical networks (XG-PON): Transmission convergence (TC) layer specification. ITU-T. January 2014.
- [5] ITU-T G.989. 40-Gigabit-capable passive optical networks (NG-PON2): Definitions, abbreviations and acronyms. ITU-T. October 2015.
- [6] Calix. Next-Generation PON: Eliminating Physical constraints from the access network, 2017, https://www.calix.com/content/dam/calix/marketing-documents/public/Calix_Technical_Brief_NG-PON2.pdf
- [7] ITU-T G.694.1. Series G: Transmission Systems and Media, Digital Systems and Networks: Transmission media and optical systems characteristics – Characteristics of optical systems. Spectral grids for WDM applications: DWDM frequency grid. 2012.

Problematika smerovania v Cloud MANET

¹Natalia KURKINA, ²Ján PAPAJ, ³Maroš BAUMGARTNER

^{1,2,3} Katedra elektroniky a multimediálnych telekomunikácií, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

¹natalia.kurkina@tuke.sk, ²jan.papaj@tuke.sk, ³maros.baumgartner@tuke.sk

Abstract — Mobilná ad hoc sieť (MANET) je populárna sieť, ktorá poskytuje viacero núdzových služieb. Integrácia MANET s cloudovými technológiami má za cieľ dosiahnuť vysokú rýchlosť prenosu dát pri súčasnom znížení latencie a nákladov. Smerovacie protokoly navrhnuté pre Cloud MANET musia byť flexibilné a schopné prispôbiť sa dynamickým charakteristikám týchto sietí. V súčasnosti je používanie algoritmov strojového učenia (ML) v bezdrôtových sieťach čoraz populárnejšie, vrátane MANET sietí, ktoré majú výhodu vyberať z možných predpovedí založených na údajoch. Táto štúdia zahŕňa hlavné procesy prvého kroku vytvárania Cloud MANET a zber informácií o sieti, na základe ktorých v budúcich štúdiách budú vyriešené problémy smerovania v takýchto sieťach.

Keľúčové slová — Cloud MANET, ad hoc sieť, strojové učenie, smerovací protokol, zhlukovanie

I. ÚVOD

MANET je mobilná ad hoc sieť. Rozvoj tohto typu sietí sa začal v 70 rokoch 19 storočia a teraz získavajú čoraz väčšiu obľubu. Siete MANET nemajú pevnú infraštruktúru, majú decentralizovaný charakter a vzhľadom na vysokú mobilitu uzlov majú neustále sa meniacu topológiu. Každý uzol v sieti funguje ako smerovač a je schopný preposielať informácie iným uzlom. Využitie takýchto sietí môže byť užitočné v oblastiach, kde neexistuje pevná infraštruktúra alebo bola zničená, napríklad v oblastiach postihnutých zemetraseniami [1].

V posledných rokoch sa počet aplikácií pre takéto siete zvýšil. V závislosti od funkcie a oblasti využitia ich možno rozdeliť na: automobilové ad hoc siete (Vehicular ad hoc networks, VANETs, - mobilná ad hoc sieť spájajúca pohybujúce sa vozidlá), cloudová mobilná ad hoc sieť (Cloud MANET, spája dve oblasti: mobilné ad hoc siete a cloudové riešenia), ad hoc siete pre smartfóny (Smart phone ad hoc networks, SPANs, spája smartfóny do jednej siete bez použitia mobilných sietí, bezdrôtových sietí alebo tradičných sietí), internetové mobilné ad hoc siete (Internet based mobile ad hoc networks, iMANETs, kombinujú mobilné ad hoc siete a Internet. Tento typ siete možno považovať za variant cloudových MANET, kde Internet funguje ako cloud) [2].

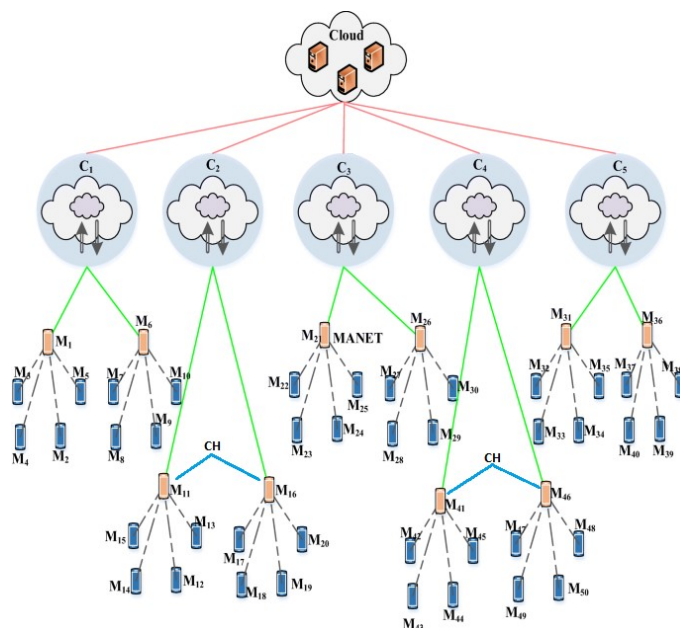
Tento článok sa zaoberá charakteristikami cloudových MANET, ich vlastnosťami, možnosťami využitia strojového učenia na zlepšenie smerovania a tým, s akými problémami sa v tomto prípade môžeme stretnúť.

II. CHARAKTERISTIKY A PROBLÉMY CLOUD MANET

Ako už bolo spomenuté, cloudové MANET spájajú dve oblasti – mobilne ad hoc siete a cloudové technológie. Hlavným cieľom takejto kombinácie bolo dosiahnuť vysokú rýchlosť prenosu dát a znížiť latenciu. Obrázok 1 reprezentuje typicky diagram takýchto sietí [3].

Zvyčajne MANET pozostáva z mobilných uzlov, ktoré sú navzájom spojené a spoločne sa podieľajú na smerovaní dát v sieti. V Cloud MANET okrem mobilných uzlov v sieti sa na rozhodovaní a presmerovaní prevádzky podieľa aj cloud. Všetky mobilné uzly v sieti sú zoskupené do klastrov. Každý uzol môže súčasne patriť len do jedného klastra. V každom klastru je vybratý jeden uzol ako hlava klastra (cluster head, CH). Zvyčajne je zodpovedný za pripojenie ku cloudu. Na prenos informácií z jedného klastra do druhého sa informácie najskôr odošlú do cloudu, potom cloud odošle údaje do zodpovedajúceho CH druhého klastra, ktorý ich presmeruje do cieľového uzla [4].

Pri kombinácii MANET s cloudom však vznikajú určité problémy. Tieto problémy sú nasledovné [4]:



Obr. 1 Cloud MANET sieť [4]

1. Časté zmeny v topológii siete vedú k častým stratám komunikácie medzi mobilnými uzlami a potrebe aktualizovať trasy v reálnom čase.
2. Simultánne prerušenia spojenia medzi uzlami spôsobujú oneskorenia a stratu paketov pri prenose dát.
3. Ťažkosť pri výbere optimálneho uzla, ktorý by mal spájať MANET s cloudom, čo môže viesť k strate prístupu do cloudu.
4. Preťaženie v CH uzloch zvyšuje stratu paketov, čo vedie k slabému výkonu systému.

Zhlukovanie/klastrovanie uzlov v takejto sieti je nevyhnutnosťou, čo uľahčuje komunikáciu medzi uzlami a zjednodušuje proces prenosu dát. Výber vhodného klastrovacieho algoritmu a definovanie hlavy klastra je jednou z hlavných úloh pri vývoji cloudových MANET. Hlava klastra často funguje aj ako uzol na komunikáciu s cloudom. Nie je to však podmienkou a na komunikáciu s cloudom je možné vybrať samostatný uzol v každom klastri [3].

Kvalita spojenia a mobilita sú hlavnými problémami smerovania v MANET kvôli častému odpájaniu a potrebe aktualizovať trasy v reálnom čase. Preto je druhou hlavnou úlohou výber optimálneho smerovacieho algoritmu.

V poslednej dobe sa vedci čoraz viac obracajú k možnostiam využitia strojového učenia na vyriešenie týchto problémov. Algoritmy strojového učenia zlepšujú procesy klastrovania a smerovania v sieti, ale stále neexistuje jediné riešenie, ktoré by vyriešilo všetky problémy. Preto je výskum v tejto oblasti stále aktuálny a potrebný.

III. STROJOVÉ UČENIE V MANET

Vďaka výpočtovému výkonu dostupnému v poslednej dobe výrazne vzrástol záujem o aplikáciu strojového učenia v rôznych oblastiach. Oblasť mobilných ad hoc sietí nie je výnimkou a vedci v poslednom čase aktívne využívajú metódy strojového učenia na zlepšenie existujúcich riešení alebo vývoj nových riešení.

Existuje mnoho rôznych algoritmov strojového učenia na zlepšenie rozhodovania zariadenia. Väčšina z nich je založená na spracovaní údajov z už existujúceho súboru údajov (dataset), ktorý už môže byť označený (učenie s učiteľom) alebo neoznačený (učenie bez učiteľa). Existujú aj algoritmy, ktoré nepoužívajú existujúce súbory údajov (reinforcement learning). Takéto algoritmy sú trénované na základe hodnotení, ktoré sú pre konkrétnu akciu prijaté z externého prostredia (napríklad, Q-learning) [5].

V sieťach MANET sa často používa strojové učenie na [5]:

- predpovedanie pohybu uzlov, čo pomáha zlepšiť smerovanie,
- odhalenie škodlivých alebo kompromitovaných uzlov, čo pomáha zvýšiť bezpečnosť v sieti,
- predpovedanie času životaschopnosti spojenia, čo pomáha vopred vypočítať novú trasu a znížiť tak oneskorenie v sieti,

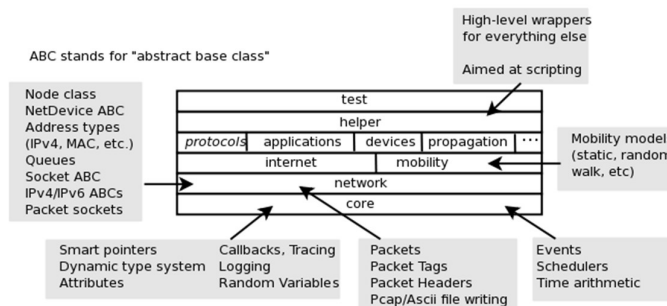
- zníženie spotreby energie a zvýšenie doby prevádzky siete.

V každom projekte založenom na strojovom učení je prvým krokom pred vývojom akéhokoľvek modelu zhromažďovania údajov, ich spracovanie a usporiadanie podľa požiadaviek modelu strojového učenia. Navyše v praxi nie je jednoduché vytvoriť označenú množinu údajov, pretože je veľmi závislá od typu údajov. Voľne dostupné dátové sady pre mobilne siete neobsahujú potrebné informácie, takže jedným z problémov pred riešením úloh súvisiacich so smerovaním je problém zberu informácií o sieti.

IV. SIMULÁCIA A ZBER ÚDAJOV

Úlohou praktickej časti tohto článku je simulácia Cloud MANET siete a zber informácie o uzloch v sieti. Tieto informácie budú použité na riešenie problémov zhlukovania a smerovania v budúcich výskumoch.

Pre simulovanie Cloud MANET bolo použité prostredie simulátora NS-3. NS-3 je softvér s otvoreným zdrojovým kódom. Je to simulátor diskretných udalostí, ktorý využíva akýkoľvek druh siete, a ktorý je realizovaný v C++ programovacím jazyku. NS-3 je implementovaný ako knižnica modulov C++, ktorú je možné povoliť/zakázať, upraviť alebo vytvoriť úplne od začiatku. Z týchto modulov bolo implementovaných mnoho ďalších modulov na modelovanie rôznych sieťových modelov, od rôznych tried smerovacích protokolov až po rôzne topológie siete. Na obr. 2 je znázornená organizácia softvérových modulov NS-3. NS-3 neobsahuje žiadne nástroje grafického používateľského rozhrania (GUI), preto pre používateľov, ktorí nie sú oboznámení s rozhraním príkazového riadka (CLI) môže byť dosť zložité. NS-3 tiež podporuje písanie simulačných skriptov v programovacom jazyku Python [6].



Obr. 2 Organizácia softvérových modulov NS3 [6].

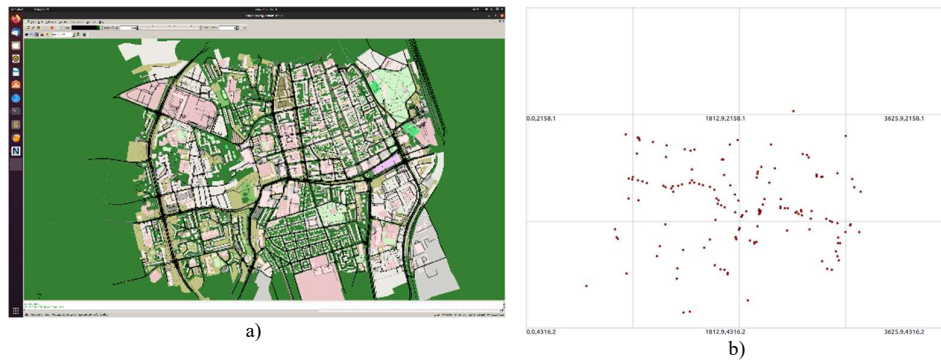
Na simuláciu pohybu uzlov v sieti, čo najbližšie k realite bol použitý ďalší simulátor SUMO. Simulation of Urban Mobility (SUMO) je multimodálny program na detailnú simuláciu dopravnej premávky s otvoreným kódom, ktorý umožní simulovať pohyb zariadení na špecifikovanej sieti ciest [7].

V našom príklade bola na simuláciu pohybu uzlov zvolená mapa centra mesta Košice, Slovenská republika. Pomocou SUMO bola urobená simulácia pohybu 150 uzlov a následne bola táto simulácia exportovaná do NS-3. Na obrázku 3 môžete vidieť príklady simulácie v SUMO a výsledok exportu do NS-3. Parametre simulácie sú uvedené v Tabuľke 1.

Tabuľka 1
 Parametre simulácie

Parameter	Hodnota
Simulation Area	3626 x 4316 m ²
Simulation Time	100 s
Network Settings	
Number of MANET users	150
Moving Speed Minimum	0 m/s
Moving Speed Maximum	20 m/s
Number of cloud nodes	1
Transmission range of MANET node	25–250 m
Channel type	Wireless

V dôsledku simulácie bol vytvorený súbor s hodnotami atribútov uzlov v sieti pre ďalšiu analýzu a aplikáciu. Pôvodný formát tohto súboru bol v ťažko čitateľnej forme, takže na jeho konverziu bol napísaný skript v programovacom jazyku Python. Príklady riadkov z tohto súboru a skonvertovaného súboru je možné vidieť na obrázku 4.



Obr. 3 Príklad simulácie v SUMO (a) a výsledok exportu do NS-3 (b).

```
<value path="/$ns3:NodeListPriv/NodeList/0/$ns3:Node/DeviceList/0/$ns3:WifiNetDevice/Phy/$ns3:VansifiPhy/Frequency" value="5860"/>
<value path="/$ns3:NodeListPriv/NodeList/0/$ns3:Node/DeviceList/0/$ns3:WifiNetDevice/Phy/$ns3:VansifiPhy/ChannelWidth" value="10"/>
<value path="/$ns3:NodeListPriv/NodeList/0/$ns3:Node/DeviceList/0/$ns3:WifiNetDevice/Phy/$ns3:VansifiPhy/ChannelNumber" value="172"/>
<value path="/$ns3:NodeListPriv/NodeList/0/$ns3:Node/DeviceList/0/$ns3:WifiNetDevice/Phy/$ns3:VansifiPhy/RxSensitivity" value="-101"/>
<value path="/$ns3:NodeListPriv/NodeList/0/$ns3:Node/DeviceList/0/$ns3:WifiNetDevice/Phy/$ns3:VansifiPhy/CcaEdThreshold" value="-62"/>
<value path="/$ns3:NodeListPriv/NodeList/0/$ns3:Node/DeviceList/0/$ns3:WifiNetDevice/Phy/$ns3:VansifiPhy/TxDgain" value="0"/>
<value path="/$ns3:NodeListPriv/NodeList/0/$ns3:Node/DeviceList/0/$ns3:WifiNetDevice/Phy/$ns3:VansifiPhy/RxDgain" value="0"/>
<value path="/$ns3:NodeListPriv/NodeList/0/$ns3:Node/DeviceList/0/$ns3:WifiNetDevice/Phy/$ns3:VansifiPhy/TxPowerLevel" value="1"/>
<value path="/$ns3:NodeListPriv/NodeList/0/$ns3:Node/DeviceList/0/$ns3:WifiNetDevice/Phy/$ns3:VansifiPhy/TxPowerEnd" value="7.5"/>
<value path="/$ns3:NodeListPriv/NodeList/0/$ns3:Node/DeviceList/0/$ns3:WifiNetDevice/Phy/$ns3:VansifiPhy/TxPowerStart" value="7.5"/>
<value path="/$ns3:NodeListPriv/NodeList/0/$ns3:Node/DeviceList/0/$ns3:WifiNetDevice/Phy/$ns3:VansifiPhy/ChannelSwitchDelay" value="+250000ns"/>
<value path="/$ns3:NodeListPriv/NodeList/0/$ns3:Node/DeviceList/0/$ns3:WifiNetDevice/Phy/$ns3:VansifiPhy/Antennas" value="1"/>
```

a)

NodeID	Frequency	ChannelWidth	ChannelNumber	RxSensitivity	CcaEdTh
0	5860	10	172	-101	-62
1	5860	10	172	-101	-62
2	5860	10	172	-101	-62
4	5860	10	172	-101	-62
5	5860	10	172	-101	-62
6	5860	10	172	-101	-62
7	5860	10	172	-101	-62

b)

Obr. 4 Príklady riadkov z pôvodného súboru (a) a po úprave (b).

V. ZAVER

Cieľom tejto práce bolo vytvoriť model siete Cloud MANET, ktorý by bolo možné použiť ďalej vo výskumných prácach pri návrhu optimálneho protokolu smerovania s použitím strojového učenia. Cieľom práce bol tiež zber hodnôt atribútov, ktoré by potom bolo možné využiť pri vytváraní klastrov v Cloud MANET a výbere CH. Na tieto účely boli použité dva simulačné programy – NS-3 a SUMO. Na úpravu vytvoreného súboru bol napísaný skript v programovacom jazyku Python.

Budúcim krokom vo výskume smerovania v Cloud MANET sieťach bude navrhnutie optimálneho algoritmu klasterizácie a výberu hlavy klastra(cluster head) s použitím hodnôt zo zostaveného súboru.

POĎAKOVANIE

Táto publikácia vznikla vďaka podpore v rámci Operačného programu Integrovaná infraštruktúra pre projekt: Výskum v sieti SANET a možnosti jej ďalšieho využitia a rozvoja, kód ITMS: NFP313010W988, spolufinancovaný zo zdrojov Európskeho fondu regionálneho rozvoja.

REFERENCIE

- [1] Sakthidasan Sankaran, K., Vasudevan, N., Devabalaji, K.R., Babu, T.S., Alhelou, H.H. & Yuvaraj, T. 2021, "A Recurrent Reward Based Learning Technique for Secure Neighbor Selection in Mobile AD-HOC Networks", *IEEE Access*, vol. 9, pp. 21735-21745.
- [2] S. K. Singh and J. Prakash, "Energy efficiency and load balancing in manet: A survey," in 2020 6th International Conference on Advanced Computing and Communication Systems, ICACCS 2020, 2020, pp. 832–837.
- [3] Alghamdi S.A., "Novel trust-aware intrusion detection and prevention system for 5g manet-cloud," *International Journal of Information Security*, 2021.
- [4] Alghamdi S.A., "Three-Tier Architecture Supporting QoS Multimedia Routing in Cloud-Assisted MANET with 5G Communication (TCM5G).", *Mobile Networks and Applications*, vol. 25, no. 6, 2020, pp. 2206-2225.
- [5] Singh K. K., A. Singh, K. Cengiz and D. Le, "Machine Learning and Cognitive Computing for Mobile Communications and Wireless Networks." 2020.
- [6] NS-3 Network Simulator Manual, NS-3 Consortium, 2021.
- [7] Dian Khumara, M. A., L. Fauziyyah, and P. Kristalina, "Estimation of urban traffic state using simulation of urban mobility(SUMO) to optimize intelligent transport system in smart city". *IES-ETA 2018 - proceedings*; 2019. 163 p.

Využitie evolučného ANFIS-u pre navigačné účely

Ján VAŠČÁK

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

jan.vascak@tuke.sk

Abstrakt – Existuje veľké množstvo metód na hľadanie cesty medzi prekážkami. Cieľom tohto článku je napodobiť spôsob uvažovania blízky človeku. Za týmto účelom pre potreby navigácie využívame produkčné pravidlá. Avšak ich návrh nie je triviálny a snažíme sa túto úlohu riešiť pomocou siete typu evolučný ANFIS, kde pre účely samotnej adaptácie parametrov bázy znalostí využívame evolučný algoritmus ako náhradu za spätné šírenie chyby. Perspektívnosť tohto prístupu bola overená pomocou sady simulačných experimentov pre potreby navigácie robota.

Kľúčové slová – evolučný ANFIS, mobilný robot, navigácia

I. ÚVOD

Autonómna navigácia bezpilotných vozidiel, tzv. UGV (unmanned ground vehicles), vyžaduje aktualizované a podrobné informácie o prostredí, v ktorom sa nachádzajú. Informácie sú získavané palubnými senzormi. Na trojrozmerné 3D snímanie sa zväčša využívajú laserové diaľkové senzory. Tie vytvárajú takzvané *mračná bodov* (*point clouds*) v priestore, v ktorom sa robot pohybuje. Tieto typy senzorov sú pre takéto vozidlá využívané napríklad na navigáciu v teréne alebo na prieskum planét, či vyhladávanie v meste alebo záchranu a poľnohospodárske účely [1], [2].

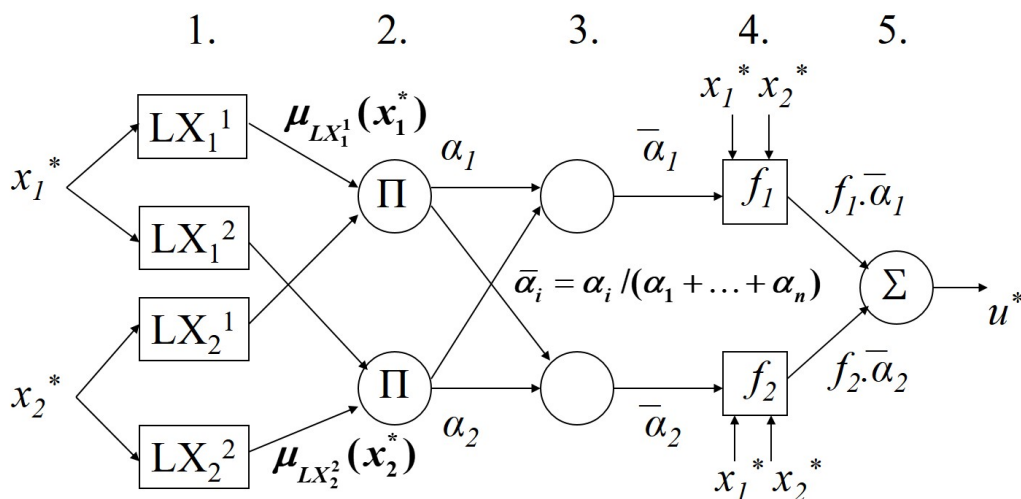
Spracovanie vonkajšieho prostredia pomocou 3D laserového skenera sa veľmi líši od obvyčajného skenovania objektov, pretože sa údaje zisťujú vo viacerých smeroch a na rôzne vzdialenosti. Prírodnú klasifikáciu terénu je možné vykonať výpočtom charakteristík, ktoré zachytávajú jednotlivé priestorové rozloženie 3D bodov, vid' [3], [4]. V minulosti bolo získavanie 3D dát realizované pomocou takzvanej „stop-and-go“ metódy, ale momentálne je ho možné vykonať pomocou simultánnej lokalizácie a mapovania (SLAM) v kombinácii s údajmi z odometrie alebo využitím inerciálnej navigačnej jednotky [5], [6].

Usporiadané 3D mračno bodov je reprezentáciou okolia, ktorú je možné taktiež použiť pre navigáciu. V tejto súvislosti sú tzv. *výškové mapy* (*elevation maps*) kompaktným spôsobom zobrazenia 3D mračien bodov. Rozšírením tohto prístupu sú *fuzzy výškové mapy* (*Fuzzy Elevation maps* – FEM), ktoré dokážu s využitím tzv. *fuzzy interpolácie* pracovať aj s chybami a nedokonalosťou senzorov. Fuzzy interpolácia však môže priniesť nespoľahlivé riešenie v oblastiach s malým počtom dát a preto je potrebné pre každú FEM použiť tzv. *fuzzy masku spoľahlivosti* (*Fuzzy Reliability Mask* – FRM), [7], [8].

Navigáciu v teréne je možné vykonávať pomocou sledovania vzdialených cieľových bodov s danými geografickými súradnicami. Tieto body na trase môžu byť plánované pomocou 2D máp alebo leteckých snímok a predstavujú trajektóriu, po ktorej by sa vozidlo malo pohybovať. Pri navigácii sa však robot musí vyhýbať miestnym prekážkam. Na riešenie tohto problému využívame v tejto práci lidar.

II. POPIS PROSTRIEDKOV A ICH MODIFIKÁCIA

ANFIS, ako skratka *Adaptive Network-based Fuzzy Inference System*, je model neurónovej, presnejšie povedané adaptívnej, siete, pomocou ktorej je realizovaný fuzzy regulátor typu TSK (Takagi-Sugeno-Kang). Hlavnou výhodou takéhoto riešenia je možnosť vytvoriť bázu pravidiel fuzzy regulátora s využitím tréningových dát a adaptačných algoritmov, ktoré sú veľmi podobné metódam učenie využívaných neurónovými sieťami. Väčšinou sa jedná o modifikovaný prístup spätného šírenia chyby, ktorý sa už osvedčil v mnohých aplikáciách neurónových sietí. Hlavným cieľom ANFIS-u je teda optimalizovať parametre ekvivalentného fuzzy regulátora pomocou súborov vstupno-výstupných dát. Parameter optimalizácie sa nastavuje tak, aby chybová hodnota



Obr. 1 Model siete ANFIS.

medzi skutočným výstupom a cieľovým výstupom bola minimálna. Inými slovami, jedná sa o tzv. kontrolované učenie.

Základná štruktúra siete ANFIS pozostáva v zásade z piatich vrstiev, ako je vidieť na Obr. 1. Prvá vrstva predstavuje výpočet tzv. kompatibility, t. j. príslušnosti vstupnej hodnoty k danej slovnej hodnote, ktorá je vyjadrená v podobe stupňa príslušnosti. Druhá vrstva agreguje, čiže prepája vstupy, a tým vytvára predpokladovú časť pravidla, ktorej miera platnosti je vyjadrená v podobe tzv. sily pravidla. Tretia vrstva normalizuje sily jednotlivých pravidiel tak, aby ich súčet bol vždy rovný 1. Štvrtá vrstva vyčíslí uje dôsledkovú časť pravidla, ktorá je vo všeobecnosti funkciou výstupnej hodnoty závislej od vstupov. Napokon piata vrstva vykonáva tzv. akumuláciu, čiže spája výstupy všetkých pravidiel s nenulovou silou do spoločnej výstupnej hodnoty.

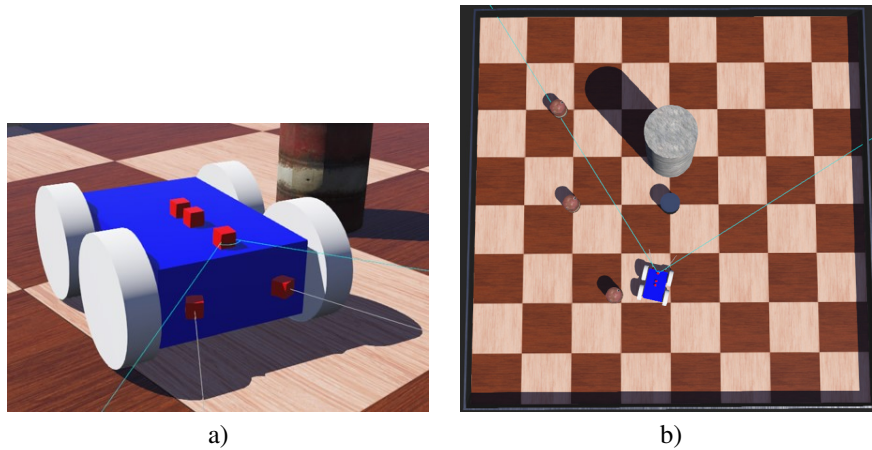
Existuje však celá rada modifikácií tohto základného prístupu. V našom prípade jeho štruktúra bude zredukovaná na štyri vrstvy. Prvá vrstva je tvorená funkciami príslušnosti, ktoré sú generované zo vstupov a z jedného vstupu vznikajú dve takéto funkcie. Z funkcií príslušnosti sú tvorené pravidlá kombinovaním funkcií príslušnosti tak, že z prvej polovice funkcií každú jednu kombinujeme s funkciami z druhej polovice spôsobom, že z tejto polovice vyberáme každú jednu funkciu zaradom a následne priradíme jednotlivým funkciám z prvej polovice. V ďalšom kroku normalizáciu priamo spájame s vyčísl'ovaním dôsledkovej časti, ktoré realizujeme pomocou lineárnej regresie.

Ďalším predmetom úpravy môže byť samotný adaptačný mechanizmus, kde sa pokúsime osvedčené spätné šírenie chyby zameniť za evolučný algoritmus. Vytvoríme populáciu sietí typu ANFIS s rôznymi nastaveniami funkcií príslušnosti vstupov a parametrov výstupných funkcií. Samotné pravidlá meniť nebudeme. Takýmto spôsobom sa pokúsime zamedziť riziku konvergencie riešenia do lokálneho extrému a lepšie prehľadať priestor možných riešení. Zároveň môžeme ľahšie zdefinovať obmedzujúce podmienky, ktoré sa týkajú tvaru a vzájomného postavenia parametrov funkcií príslušnosti tak, aby lepšie zodpovedali ľudskej predstave ich tvaru ako aj riešeniu, ak by ho navrhoval človek.

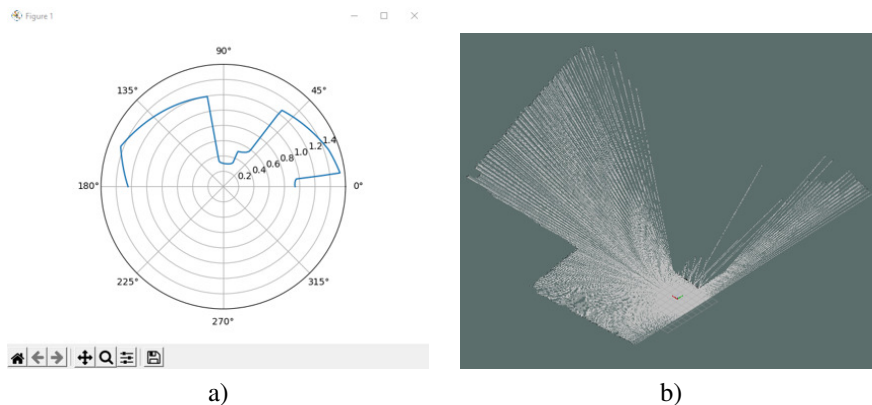
III. NÁVRH RIEŠENIA

Pre potreby experimentovania sme využili robotický simulátor Webots, v ktorom sme vytvorili virtuálneho robota, ktorý obsahuje 4 kolesá, 2 ultrazvukové diaľkomery, 1 lidar, 1 GPS prijímač a 1 rádiový kompas, vid' Obr. 2a. Svojimi vlastnosťami sa podobá na známy mobilný robot Pioneer 3-AT. Kompas udáva smer robota, GPS prijímač jeho polohu a snímače otáčok kolies je možné využiť pre odometriu, ak by spomenuté snímače nemohli byť z nejakého dôvodu využité, čo sme simulovali v podobe náhodných porúch. Podobne sme vytvorili model simulačného prostredia, kde môže rozmiestňovať prekážky rôznych tvarov a rozmerov, vid' Obr. 2b.

Úlohou v ďalšom kroku bolo pomocou lidararu vytvoriť mračno bodov, prostredníctvom ktorého by sme boli schopní vizualizovať prostredie okolia robota s prekážkami. Za týmto účelom sa použila knižnica *PyVista* pre programovací jazyk *Python*, pomocou ktorej sme vizualizovali mračno bodov. Na Obr. 3 je možné pre porovnanie vidieť, ako zobrazuje prekážku z radarového grafu s údajmi o vzdialenostiach (Obr. 3a) mračno bodov, vid' Obr. 3b.



Obr. 2 Model a) experimentálneho virtuálneho robota a b) prostredia.



Obr. 3 Vizualizácia videnia prekážky robotom pomocou a) radarového grafu a b) pomocou mračna bodov.

IV. EXPERIMENTY A VÝSLEDKY

Po vykonaní adaptácie s využitím evolučného ANFIS-u bola vykonaná rada simulácií v rôznych prostrediach s rôznym počtom prekážok. Vstupom do riadiaceho systému boli údaje zo senzorov, konkrétne súradnice x, y z lidar, ďalej súradnice robota v priestore a uhol, ktorý mal daný robot v priestore. Počas experimentov sme skúmali rozdiel vo výsledkoch medzi rôznymi hodnotami parametra mutácie a učenia. Na Obr. 4 môžeme vidieť najlepší dosiahnutý výsledok, keď pri hodnote parametra učenia 0,2 a parametra mutácie 0,2 dosiahla chyba riadenia najnižšiu hodnotu už po 500 generáciách.

V. ZÁVER

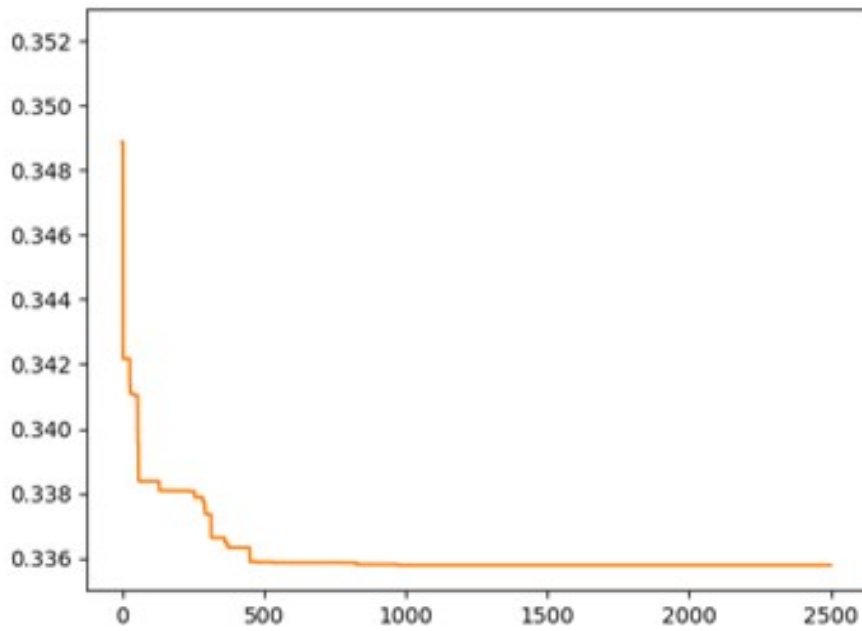
Táto práca je iba počiatočnou štúdiou využiteľnosti siete typu evolučný ANFIS, kde sme nahradili modifikovanú metódu spätného šírenia chyby evolučným algoritmom. Vzhľadom na rozmanitosť typov evolučných algoritmov bude ďalším cieľom vyskúšať ich základné typy a navrhnúť vhodnú úpravu evolučného prístupu pri adaptácii sietí ANFIS. Ďalším krokom bude automatizácia získavania dát z mračna bodov, ktoré by boli ďalej posielané a vytvorili by tak fuzzy mapu s cieľom vedieť pracovať s neurčitými dátami.

POĎAKOVANIE

Táto publikácia je výstupom projektu ENISaC – Edge-enabled inteligentné snímanie a výpočty (APVV-20-0247), podporovaného Agentúrou na podporu vedy a výskumu (APVV).

LITERATÚRA

- [1] Larson, J.; Trivedi, M.; Bruch, M. "Off-road terrain traversability analysis and hazard avoidance for UGVs", In Proceedings of the IEEE Intelligent Vehicles Symposium (IV), Baden-Baden, Germany, 6 February 2011; pp. 1–7.
- [2] Weiss, U.; Biber, P. "Plant detection and mapping for agricultural robots using a 3D LIDAR sensor", Robot. Auton. Syst. 2011, 59, pp. 265–273.
- [3] Lalonde, J.F.; Vandapel, N.; Huber, D.F.; Hebert, M. "Natural terrain classification using three-dimensional lidar data for ground robot mobility", J. Field Robot. 2006, 23, pp. 839–861.



Obr. 4 Priebeh učenia systému (parameter učenia = 0,2; parameter mutácie = 0,2).

- [4] Plaza-Leiva, V.; Gomez-Ruiz, J.A.; Mandow, A.; García-Cerezo, A. "Voxel-Based Neighborhood for Spatial Shape Pattern Classification of Lidar Point Clouds with Supervised Learning", *Sensors* 2017, 17, 594.
- [5] Martínez, J.L.; Mandow, A.; Reina, A.J.; Cantador, T.J.; Morales, J.; García-Cerezo, A. "Navigability analysis of natural terrains with fuzzy elevation maps from ground-based 3D range scans", In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Tokyo, Japan, 3–7 November 2013; pp. 1576–1581.
- [6] Mandow, A.; Cantador, T.J.; Reina, A.J.; Martínez, J.L.; Morales, J.; García-Cerezo, A. "Building Fuzzy Elevation Maps from a Ground-Based 3D Laser Scan for Outdoor Mobile Robots", In *Advances in Intelligent Systems and Computing*; Springer: Basel, Switzerland, 2016; Volume 417, pp. 29–41.
- [7] Lacroix, S.; Mallet, A.; Bonnafous, D.; Bauzil, G.; Fleury, S.; Herrb, M.; Chatila, R. "Autonomous rover navigation on unknown terrains: Functions and integration", *Int. J. Robot. Res.* 2002, 21, pp. 917–942.
- [8] Seraji, H. "SmartNav: A rule-free fuzzy approach to rover navigation", *J. Field Robot.* 2005, 22, pp. 795–808.

Detekcia sarkazmu na sociálnych sieťach

¹Martin Sedliak, ²Martina Szabóová

^{1,2}Katedra kybernetiky a umelej inteligencie, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

¹martin.sedliak@student.tuke.sk ²martina.szaboova@tuke.sk

Abstrakt – Článok sa zaoberá detekciou sarkazmu použitím rôznych algoritmov strojového učenia. Na začiatku podáva prehľad problematiky detekcie sarkazmu. Využitím analýzy doterajších prístupov riešenia problému detekcie sarkazmu sú zvolené vhodné metódy strojového učenia: metóda podporných vektorov, logistická regresia, Naivný Bayesov klasifikátor a konvolučná neurónová sieť. V závere sú vyhodnotené navrhnuté riešenia.

Kľúčové slová – detekcia sarkazmu, konvolučné neurónové siete, metóda podporných vektorov, Naivný Bayesov klasifikátor

I. ÚVOD

Sarkazmus je forma verbálnej irónie, ku ktorej dochádza, keď existuje rozpor medzi doslovným a zamýšľaným významom výroku. Prostredníctvom tohto rozporu rečník vyjadruje svoj postoj k predchádzajúcemu návrhu, často vo forme povrchného pohrdania alebo krivdy. Uveďme si ako príklad vetu: “Dobre si to spravil.“ Bez kontextu a všeobecného prehľadu nevieme rozpoznať, či táto veta vyjadruje sarkazmus. Sarkazmus je všadeprítomný na sociálnom webe a vzhľadom na svoju povahu, môže veľmi narúšať výpočtové systémy, ktoré využívajú tieto údaje na vykonávanie úloh, ako je analýza sentimentu, získavanie názorov, profilovanie autorov a zisťovanie obťažovania.

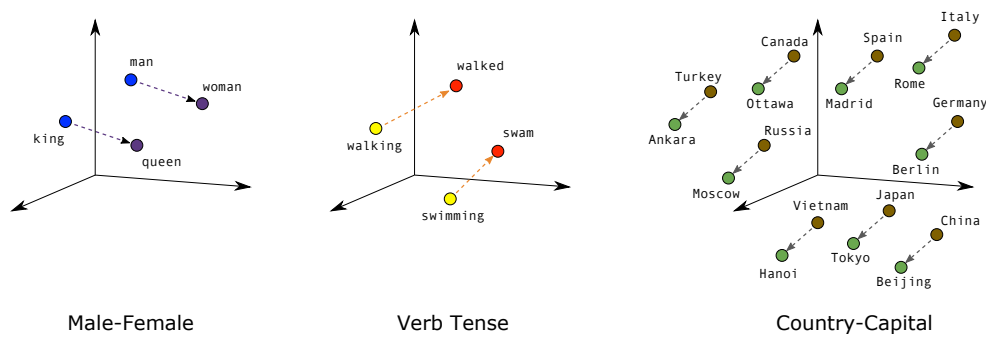
Jedným z dôvodov, prečo je téma detekcie sarkazmu aktuálna je fakt, že sarkazmus je sofistikovaná forma reči, v ktorej sa povrchový význam líši od implikovaného významu. Táto forma reči vyjadruje myšlienku implicitne, takže je ťažké odhaliť sarkazmus v prejave. Keďže rétorika v sarkazme závisí od kontextu, je ťažké poznať zámery rečníka, pokiaľ nedostaneme priamo od neho poznámku, či sa jedná o sarkazmus alebo nie. Rozpoznanie sarkazmu je rozhodujúce pre pochopenie skutočného sentimentu a významu prejavu človeka. Ťažkosti s rozpoznávaním sarkazmu spôsobujú nedorozumenia v každodennej komunikácii. Spôsobuje problémy v mnohých systémoch spracovania prirodzeného jazyka, ako napríklad analýza sentimentu, získavanie názorov, profilovanie autorov a zisťovanie obťažovania. Preto je nevyhnutné vyvinúť automatické detektory sarkazmu, ktoré pomôžu pochopiť implicitný význam sarkastickej odpovede [1], [2], [3].

Na vytvorenie automatických detektorov sarkazmu je potrebné využiť metódy strojového učenia a umelej inteligencie. Aby sme tieto metódy mohli použiť, je potrebné okrem ich naprogramovania aj ich tréning a testovanie, na ktoré potrebujeme vhodné dáta. Tieto dáta sú vo forme krátkych textov nazývaných “príspevky“ zo sociálnych sietí ako Reddit alebo Twitter. Takýto súbor dát pozostáva okrem krátkeho textu aj z označenia od autora, či sa jedná o sarkastický text alebo nie.

II. ANALÝZA DOTERAJŠÍCH PRÍSTUPOV K PROBLÉMU DETEKIE SARKAZMU

Prvý prístup, s ktorým sme sa stretli je založený na využití takzvaných predtrénovaných *word embeddings*. Tento pojem znamená použitie reprezentácie slov na analýzu textu vo forme vektorov. Slová s podobným významom sú usporiadané bližšie k sebe ako je možné vidieť na obrázku 1 [1]. Jednou z možností ako využiť túto metódu je pomocou modelu BERT. BERT je predtrénovaný jazykový model založený na strojovom učení, tento model bol vyvinutý za účelom zlepšiť spracovanie prirodzeného jazyka. RoBERTa je replikácia modelu BERT tréňovaná na dvanásťkrát väčšom súbore dát ako BERT. RoBERTa využíva väčšiu slovnú zásobu, ktorá pomáha dosiahnuť lepšie výsledky ako BERT v rôznych nadväzujúcich úlohách [1].

Ďalší prístup, s ktorým sme sa stretli, sa zameriava na účinnosť kontextu informácie pri rozhodovaní, či sa jedná o sarkastický výrok alebo nie. Autori v [4] zvolili najprv klasické modely, ktoré boli vytvorené pomocou algoritmov strojového učenia ako logistická regresia, SVM, Naivný Bayesov klasifikátor. Tieto modely boli tréňované na klasifikáciu výrokov bez zohľadnenia ich kontextových informácií. Týmto spôsobom boli natréňované aj zložitejšie modely



Obr. 1 Word embeddings [1]

ako rekurentné neurónové siete (RNN) a neurónové siete s pamaťou (LSTM). Ďalšou zvolenou metodikou autorov bolo využitie word embeddings. Špecifickejšie sa jednalo o typ ELMo, a taktiež sentence embedding DeepMoji spolu s predtrénovanou neurónovou sieťou na zachytenie emócií vo vetách. Poslednými použitými modelmi v tomto experimente boli BERT a RoBERTa, ktoré sa používali v dvoch prípadoch. V prvom prípade sa modely trénovali bez kontextovej informácie patriacej k príspevku. V druhom prípade sa trénovali spolu s kontextovou informáciou príspevku.

V článku [3] použili autori model hlbokého učenia BiLSTM-CNN, kde CNN označuje konvolučnú neurónovú sieť, Bi znamená obojsmerná a LSTM je typ rekurentnej neurónovej siete, ktorá bola spomenutá v predošlom experimente. Pre trénovanie modelu bola realizovaná extrakcia vlastností osobitne pre anglický a hinduistický jazyk. Pre anglické slová bol využitý model GloVe word embeddings a pre hinduistické využili Hindi-SentiWordNet a TF-IDF. Následne bol model BiLSTM-CNN natrénovaný pomocou kombinácie anglických, hinduistických a pomocných vektorov.

V článku [5] využili autori typ rekurentnej neurónovej siete BiLSTM, ktorá je schopná učiť sa a zapamätať si dlhodobých závislostí bez toho, aby nastali problémy miznúceho alebo explodujúceho gradientu.

V poslednom prístupe, s ktorým sme sa stretli, autori [6] použili dva súbory dát. Prvý súbor obsahoval príspevky z Twitteru, druhý z Redditu. Oba súbory dát obsahovali kontextovú informáciu o príspevku, príspevok a jeho označenie či sa jedná o sarkazmus alebo nie. V tomto experimente autori prišli s modelom, ktorého architektúra pozostáva z dvoch častí. Prvá časť tvorí BERT. Druhá časť tvoria pooling-ové vrstvy, ktoré sú rozložené do BiLSTM a NeXtVLAD. NeXtVLAD je vylepšená verzia NetVLAD, ktorý pozostáva z modelu konvolučnej neurónovej siete s vyššou odolnosťou voči k pretrénovaniu. Na augmentáciu dát použili metódu spätného prekladu, ktorá spočíva v tom, že sa vety preložia do iného jazyka a potom späť do pôvodného jazyka so snahou dostať preformulovaný text. Pri tejto metóde použili jazyky francúzština, španielčina a holandčina.

III. IMPLEMENTÁCIA ZVOLENÝCH ALGORITMOV RIEŠENIA PROBLÉMU DETEKcie SARKAZMU

Využili sme štyri metódy strojového učenia, ktoré spadajú pod kontrolované učenie: metóda podporných vektorov (SVM), Logistická regresia (LR), Naivný Bayesov klasifikátor (NB) a konvolučná neurónová sieť (CNN). Kontrolované učenie je charakterizované konceptom učiteľa, ktorého hlavnou úlohou je poskytnúť algoritmu presné meranie jeho chyby (priamo porovnateľné s výstupnými hodnotami). Pri aktuálnych algoritmoch túto funkciu zabezpečuje trénovacia množina tvorená párami vstup a očakávaný výstup. Na základe týchto informácií môže agent opraviť svoje parametre tak, aby sa znížila veľkosť globálnej chybovej funkcie. Ak je algoritmus dostatočne flexibilný a dátové prvky sú koherentné, celková presnosť sa po každej iterácii zvyšuje a rozdiel medzi aproximačným a požadovaným výsledkom sa blíži k nule. Cieľom kontrolovaného učenia je trénovať systém, ktorý musí pracovať aj so vzorkami, ktoré predtým ešte nevidel. Preto je potrebné umožniť modelu vyvinúť schopnosť zovšeobecňovania a vyhnúť sa bežnému problému nazývanému *overfitting*, ktorý nastáva v dôsledku pretrénovania modelu alebo zložitosti modelu. Jedným z hlavných dôsledkov takéhoto problému je schopnosť správne predpovedať iba dáta, ktoré boli použité pri tréningu, zatiaľ čo chyba u dát na testovanie je vždy vysoká [2].

A. Predspracovanie dát

Dataset pochádzal zo súťaže SemEval 2022. Celkovo obsahoval 3467 príspevkov, z toho bolo 867 sarkastických a 2600 nesarkastických. Z uvedených počtov vyplýva, že dataset nebol vyvážený.

Tabuľka 1
Vyhodnotenie všetkých experimentov

	SVM	NB	Logistická regresia	Konvolučná neurónová sieť
Správnosť (bez úpravy dát)	0.74	0.74	0.74	0.74
Správnosť (s rozšírením množiny dát)	0.83	0.76	0.79	0.81
Správnosť (s orezaním množiny dát)	0.60	0.61	0.60	0.55

Ďalším krokom po stiahnutí dát, bolo ich rozdelenie na trénovaciu a testovaciu množinu. Následne sme vykonali klasické úpravy textu ako je tokenizácie, lematizácia, stemovanie a vymazanie stopových slov (z angl. Stop words). Potom sme text vektorizovali použitím TF-IDF (z angl. Term Frequency – Inverse Document Frequency) [7].

B. Experimenty bez úpravy dát

Na dáta získané z predspracovanie sme aplikovali vyššie spomenuté algoritmy strojového učenia, menovite: Naivný Bayesov klasifikátor, metóda podporných vektorov, logistická regresia a konvolučné neurónové siete. Všetky klasické modely strojového učenia ako aj náš model konvulčnej neurónovej siete dosiahli pri zadanej úlohe detekcie sarkazmu správnosť 74%. Modely dokázali klasifikovať len nesarkastické príspevky. Po analýze výsledkov sme došli k záveru, že za tento fenomén môže nevyvážený súbor dát, ktorý obsahoval 2600 nesarkastických a 867 sarkastických príspevkov.

C. Experimenty s úpravou množiny dát

Opäť sme použili predspracované dáta, na ktoré sme navyše implementovali metódu augmentácie. K tomuto kroku bolo potrebné pristúpiť, aby sme odstránili klasifikáciu modelov iba do jednej triedy. Túto metódu sme použili po analýze rôznych prístupov k detekcii sarkazmu, kde autori uviedli augmentáciu dát ako prostriedok, pre vyrovnanie počtu dát pri jednotlivých triedach. Augmentáciou dát sme dokázali zlepšiť výkon jednotlivých modelov, ktoré sme použili pri detekcii sarkazmu. Implementáciu augmentácie dát sme vyriešili dvoma metódami.

1) *Augmentácia substitúciou*: Prvou metódou bola substitúcia, ktorá spadá pod metódy založených na pridávaní spojitého alebo nespojitého šumu. Bol načítaný predtrénovaný GloVe word embeddings [8], z ktorého boli čerpané synonymá. Následne boli nastavené parametre na augmentáciu dát, a takto bolo vytvorených 1200 ďalších sarkastických príspevkov, ktoré boli pripojené k už existujúcemu súboru dát. Ďalším krokom bolo rozdelenie dát na trénovaciu a testovaciu množinu. Najlepšie hodnotenie dosiahla klasická metóda strojového učenia SVM so správnosťou 83%, na druhom mieste sa umiestnil náš model konvulčnej neurónovej siete so správnosťou 81%. Ostatné metódy strojového učenia skončili za modelom konvulčnej neurónovej siete so správnosťou menšou ako 80%.

2) *Augmentácia orezaním dát*: Druhou metódou augmentácie bolo orezanie dát. Náhodne bolo vybraných 860 sarkastických a 860 nesarkastických príspevkov, s ktorými sa pracovalo ďalej. Ďalším krokom bolo rozdelenie dát na trénovaciu a testovaciu množinu.

Najlepšie hodnotenie dosiahli klasické modely strojového učenia so správnosťou 60%. Model konvulčnej neurónovej siete dosiahol správnosť 55%.

D. Celkové výsledky experimentov

Na vyhodnotenie experimentov sme zostavili pri každom experimente kontingenčné tabuľky pre jednotlivé algoritmy a následne vyrátali presnosť (z angl. precision), návratnosť (z angl. recall), F1 skóre a správnosť (z angl. accuracy) [9]. Čiastkové výsledky boli spomenuté pri každom experimente.

S implementovanými klasickými modelmi strojového učenia a konvulčnej neurónovej siete boli vykonané tri experimenty. Dosiahnuté výsledky sú zhrnuté v tabuľke 1. Z uvedenej tabuľky je vidieť, že výsledkom prvého experimentu boli natrénované a otestované modely preučené. V druhom experimente sme pred trénovaním a testovaním použili augmentáciu dát na vygenerovanie nových sarkastických príspevkov. V tomto prípade dosiahla najlepší výsledok metóda SVM so správnosťou 83%. V poslednom experimente sme pred trénovaním a testovaním modelov dataset orezovali aby mali rovnaké zastúpenie obe triedy. Najlepší výsledok dosiahla metóda logistická regresia so správnosťou 61%. Najlepší výsledok zo všetkých experimentov teda dosiahla metóda SVM so správnosťou 83%.

IV. ZÁVER

V tomto článku sme sa zaoberali detekciou sarkazmu na príspevkoch sociálne siete Twitter. Identifikácia sarkazmu je samo o sebe náročná úloha aj pre bežného človeka. Avšak môžeme povedať, že naše modely dosiahli obstojnú správnosť a je možné pokračovať s ich vylepšením. Z vykonaných experimentov môžeme skonštatovať, že metóda SVM so správnosťou 83% sa javí ako najlepšia možnosť na identifikáciu sarkazmu. Avšak rozšírením nášho prístupu klasifikácie o nové informácie ako napríklad kontextové informácie alebo doladenie word embeddings na sarkastických príspevkoch by mohlo mať vplyv na výsledky ostatných algoritmov.

LITERATÚRA

- [1] T. Dadu and K. Pant, "Sarcasm detection using context separators in online discourse," in *Proceedings of the Second Workshop on Figurative Language Processing*. Online: Association for Computational Linguistics, Jul. 2020, pp. 51–55. [Online]. Available: <https://aclanthology.org/2020.figlang-1.6>
- [2] A. Kumar Jena, A. Sinha, and R. Agarwal, "C-net: Contextual network for sarcasm detection," in *Proceedings of the Second Workshop on Figurative Language Processing*. Online: Association for Computational Linguistics, Jul. 2020, pp. 61–66. [Online]. Available: <https://aclanthology.org/2020.figlang-1.8>
- [3] D. K. Jain, A. Kumar, and G. Garg, "Sarcasm detection in mash-up language using soft-attention based bi-directional lstm and feature-rich cnn," *Appl. Soft Comput.*, vol. 91, p. 106198, 2020.
- [4] A. Kumar Jena, A. Sinha, and R. Agarwal, "C-net: Contextual network for sarcasm detection," in *Proceedings of the Second Workshop on Figurative Language Processing*. Online: Association for Computational Linguistics, Jul. 2020, pp. 61–66. [Online]. Available: <https://aclanthology.org/2020.figlang-1.8>
- [5] A. Kumar, V. T. Narapareddy, V. Aditya Srikanth, A. Malapati, and L. B. M. Neti, "Sarcasm detection using multi-head attention based bidirectional lstm," *IEEE Access*, vol. 8, pp. 6388–6397, 2020.
- [6] H. Lee, Y. Yu, and G. Kim, "Augmenting data for sarcasm detection with unlabeled conversation context," in *Proceedings of the Second Workshop on Figurative Language Processing*. Online: Association for Computational Linguistics, Jul. 2020, pp. 12–17. [Online]. Available: <https://aclanthology.org/2020.figlang-1.2>
- [7] J. Ramos, "Using tf-idf to determine word relevance in document queries," 1999.
- [8] J. Pennington, R. Socher, and C. D. Manning, "Glove: Global vectors for word representation," in *Empirical Methods in Natural Language Processing (EMNLP)*, 2014, pp. 1532–1543. [Online]. Available: <http://www.aclweb.org/anthology/D14-1162>
- [9] C. Goutte and E. Gaussier, "A probabilistic interpretation of precision, recall and $j_i f_i / i_i$ -score, with implication for evaluation," in *Proceedings of the 27th European Conference on Advances in Information Retrieval Research*, ser. ECIR'05. Berlin, Heidelberg: Springer-Verlag, 2005, p. 345–359. [Online]. Available: https://doi.org/10.1007/978-3-540-31865-1_25

Comparing the Performance of Current Tools for Dynamic Code Analysis

¹Miriama JACKANINOVA, ²Ján HURTUK

^{1,2}Department of Computers and Informatics, Faculty of Electrical Engineering and Informatics Technical University of Košice, Slovak Republic

¹miriama.jackaninova@student.tuke.sk, ²jan.hurtuk@tuke.sk

Abstract – The aim of this paper is the familiarization with an actual state of code analysis, a description of various types of analyses with a resulting focus on the dynamic code analysis. The examples of the tools and techniques which realize the dynamic analysis are listed within this analysis. Based on an obtained knowledge, what follows is the choice of the most suitable tools for their testing and for a comparison of their performance. The available sandboxes which create the so called virtual environment are chosen as the tools of analysis. Open-source Sandboxie-Plus, shareware Shade Sandbox and Windows Sandbox, which is available within an update OS Windows, are included among the compared sandboxes. The performance of the chosen tools is compared on the basis of available technical parameters, as well as on the basis of launching five ransomware samples in the environment which they offer. The result of the testing is a finding that Sandboxie-Plus tool offers possibility of collective testing of the samples with a 100 per cent success rate of the detection of ransomware samples. It hereby reached the first position in the testing of sandbox tools of the dynamic code analysis.

Keywords – Dynamic analysis, static analysis, hybrid analysis, Machine learning, sandbox, ransomware, Sandboxie-Plus, Shade Sandbox, Windows Sandbox

I. INTRODUCTION

Today's data cost is too high to not to protect them. Just a small mistake in the program can lead to its malfunction or the loss of some information. Loss of information is often associated with financial and time losses. Thanks the enormous development of digitisation and the diffusion of smart devices, people's doubts about their security are growing. Security testing means looking for possible weaknesses in the program, which the attacker could use to his advantage. For example, taking control of significant strategic or business nodes, or capture user's data.

Therefore, the best way is to start analyzing the code already during its creation. Error detection in the early stages is often financially and time efficient. For this is adequate to use static form of analysis that examines the code without its execution.

However, in our work we focus on dynamic analysis which examines the code in its execution process. We describe a few control methods, but primarily we are focused on sandboxing. Using available sandbox tools, Sandboxie-Plus, Shade Sandbox and Windows Sandbox we perform a series of examples focussing on the possibilities that the tools offer. We execute process of dynamic code analysis with malicious ransomware samples and point out their behavior in the given tools.

The impulse for writing this work was the fact that in the recent years the ransomware attacks have been spread. The attackers are always trying to make the development of ransomware progress faster and its detection has become increasingly challenging for security creators. That is why it is necessary to test tools designed to analyse harmful samples and comparing their performance in practical way. It is important to establish a suitable tool for dynamic code analysis that will suit to user and also will be fulfilled necessary needs of the given tool, such as disk space or RAM and CPU consumption. Also, don't forget to operating system version because some tools are not supported in older versions.

The aim of this work is in addition to theoretical reference to selected sandbox tools also their remelting into the practice. There is also a concern to finding out which of them is the most user-friendly and efficient and which of them is the weakest.

II. TYPES OF CODE ANALYSES

For proper code operation is necessary to do the code analysis that tests it and points out its vulnerabilities. Currently, there are several types of analyses, which work in different way. The following subchapters are devoted to these analyses.

A. *Static analysis*

This type of analysis is considered to be the oldest type of program analysis. This analysis reviews written code without the need of program running. It mainly verify the support of the environment and tools that are used to create the program [1].

As Chess and West [2] mention, today we know three ways of static analysis applica- tion:

- Application for program optimization
- Application for program correctness
- Application in program development

We usually describe the static analysis as an automatic code inspection process, which has its pros and cons. As advantage we can concern the fact that she is going through the whole code, not just the one that can be tested later in running process. Among the disadvantage, there is for example fact that she can not accurately determine leaks in memory or program parallelism errors.

B. *Dynamic analysis*

The essence of dynamic code analysis is to investigate the application either during execution program or after its termination. Between the most common use of this analysis is the search for malware. For its most used forms we can consider a Trojan horse, a virus or even a worm, or a logic bomb [3].

As Mezquita [4] mentions, the dynamic code parsing process consists of four categories.

- data collection and preparation
- the process of program running
- gathering the necessary program parameters
- put the output data to analysis

According to Formánek [5] among the most basic dynamic technological procedures belong to:

- **Debugging** - classical method which uses a tool called a debugger. It can be used to run the code step by step and monitor program behavior.
- **Fuzzing** - this method achieves generation of unexpected entry by sending a large amount of data to the tested application.
- **Cooperation with static analysis** - usually, dynamic analysis is combined with static analysis. It's happening mainly due to an increase in the percentage of possible detected parts of the application that could make unwanted changes to the program.
- **Sandboxing** - As Ingalls [6] says, anti-virus protection is no longer sufficient against unwanted threats. Based on that so many organizations have started to use also the possibility of sandboxing. Sandbox implementation consists of using third-party software, virtual machines, embedded software or use so-called browser plug-ins. In next chapter, Sandboxes, we present three sandboxes for more details. Two as third-party software, Sandboxie-Plus and Shade Sandbox. We describe also Windows Sandbox as embedded software.

C. *Hybrid analysis*

Hybrid analysis is combination of static and dynamic analysis. From work of Roundy and Miller [7] is generally known that they are using combined algorithm to detect malware. And at the end the program is able to be as it was before the malware attack. To achieve this, they are using static technique, parsing. From dynamic techniques:

- obfuscation
- capture code properties during its execution
- self - modifying code

D. Machine learning

This is ability of computer system to acquire knowledge of the algorithms and techniques used in the code. Therefore, we can say that this is a learning process based on the experience gained from previous inputs and their results.

Based on available information, Židzik [8] classifies these types of ML (Machine learning):

- Supervised learning
- Unsupervised learning
- Combined type
- Reinforcement learning

III. SANDBOXES

The most widely used method that can provide malware detection is according to Jamalpur and his colleagues [9] sandboxing. The advantage of sandboxing is that using a sandbox that represents a certain approach running programs and files in a virtual environment, there is no physical damage, like damage of selected resources.

Analysis is a gradual process of taking the necessary steps. The process of processing malware in a sandbox environment is therefore composed of the following steps [10]:

- 1)start the object scanning process (URL or file)
- 2)sample run
- 3)the sandbox collects artifacts acquired during a specified period of time
- 4)analysis of artifacts and delivery of the verdict to the requesting system with a description "malware" or "harmless"

Among the collected artifacts we can include memory dumps, application execution logs, lists of loaded modules, network operation or changes of system files and registers.

In the next chapters we are focusing on freely available sandboxes, which are also used in our research part of work.

A. Sandboxie-Plus

Sandboxie-Plus has one feature that is different from regular sandbox tool. The feature is that Sandboxie-Plus doesn't virtualize everything. It virtualizes only the necessary resources that are required by running programs inside our sandbox. Using Sandboxie-Plus can run various processes, objects on threads, objects used for interprocessor communication such as pipes, events, mutexes or traffic lights.

On figure (Fig. 1) we can see an initial preview of the Sandboxie-Plus environment, which was introduced by the British company Sophos. Sandboxie-Plus was marketed as a freely available tool that can be easily found, download and install. It is adapted for the Windows operating system.

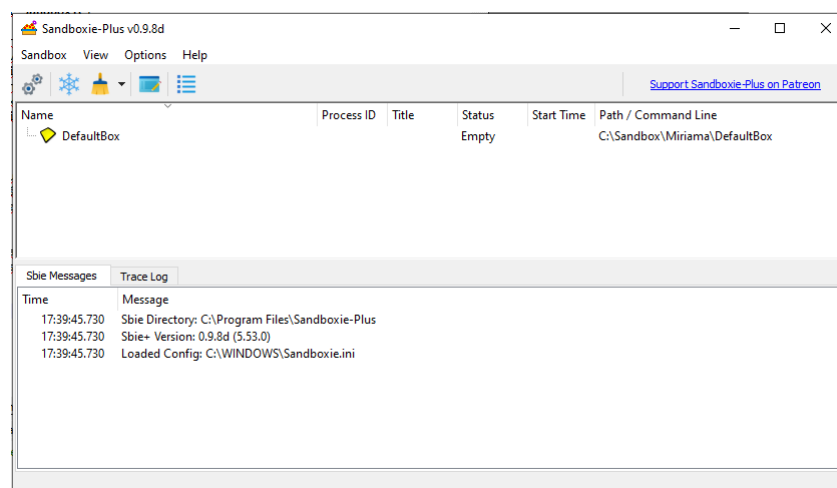


Fig. 1 An initial preview of the Sandboxie-Plus environment.

B. Shade Sandbox

Based on Phillips [11], the Shade Sandbox tool (Fig.2) is at the top of the list of best sandbox tools. This tool also represents a virtual system preventing malicious content from entering our system, developed by Eugen Balaban.

According to Hillier, the basic benefits of using Shade Sandbox [12] include:

- drag&drop function
- compatibility with antivirus programs
- low cost
- capture malware in a virtual environment
- safe internet browsing

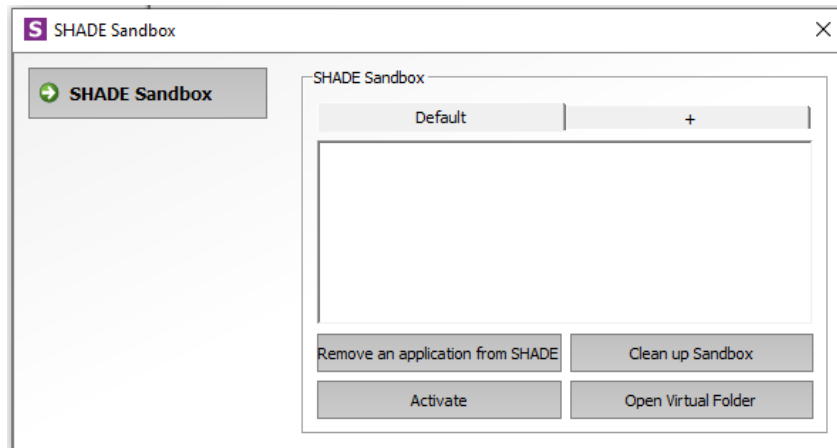


Fig. 2 Shade Sandbox - default environment.

C. Windows Sandbox

Due to the relatively frequent use of sandboxes, Microsoft [13] decided that the update for Windows 10, released in May 2019, would be contain a new feature, Windows Sandbox (Fig.3).

This sandbox is temporary because all files and applications are deleted when you close the Windows Sandbox. So every time it's started, it's available as a whole new environment. Therefore, this feature can be included among his disadvantages along with the fact that guest applications are not available for Windows Sandbox. If we want to use software or an application that is installed in the guest space, we must explicitly install it within the sandbox environment.

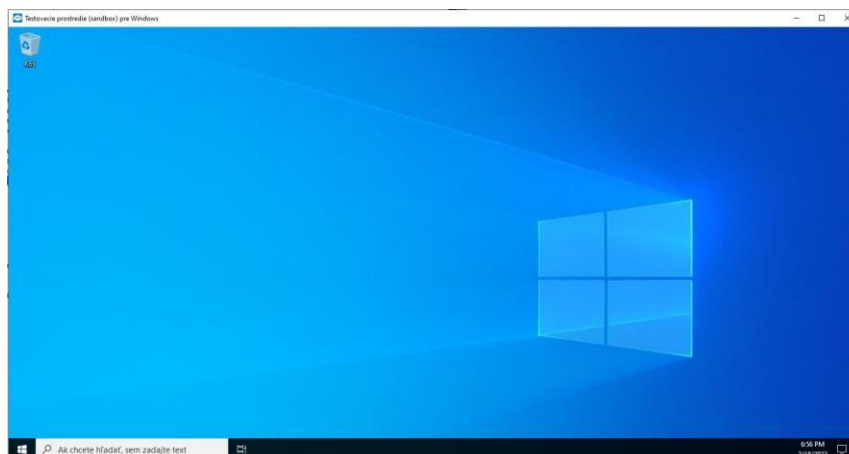


Fig. 3 Windows Sandbox - default environment.

IV. USE OF SANDBOXES IN PRACTICE

Today, sandboxes are among the most widely used tools for malware detection. As a result, it is important for us to find out which works better than others. As part of the practical demonstration, we point out the various possibilities offered by the sandbox tools. We also try to find out answers to a few important theses, which we have written in points below:

- launch internet browser in sandbox tool environment - in this point we focus on the possible user access to the Internet
- option for adding an application to the sandbox tool
- force application launch in sandbox space - exploring options for automatic or forced launch of the application or the entire folder in a sandbox environment
- sandbox vs ransomware - in this section we examine the behavior of five ransomware samples in three sandbox tools
- number of possible sandboxes
- number of possible executable samples
- RAM and CPU parameters

As part of testing sandbox vs ransomware we are using ransomware samples listed below and the launch process diagram is on Fig.4.

- **WannaCry** - According to Akbanov and his colleagues [14] WannaCry, also known as Wana Decrypt0r or WCry is one of the most widespread ransomwares. In 2017 he attacked various sectors of life such as government, healthcare or telecommunications. We classify this malware as a crypto ransomver, which encrypts data into unreadable form. Its required ransom ranges from \$300 to \$600 in bitcoins.
- **Jigsaw** - Based on Cooper [15], Jigsaw is also known as BitcoinBlackmailer, which plays its role on computers running on Windows. We consider spam emails with attachments as the most common ways in which Jigsaw spreads among its victims. The most commonly requested ransom is \$150.
- **Petya** - According to the work of Aidan and his colleagues [16], the first victims of this sample were recorded in 2016 in more than sixty countries around the world. The malicious file is distributed via emails that contain an alleged link to fill out a form. The ransom's around \$400 and it doubles every week.
- **Bad Rabbit** - Based on work of Perekalin [17], the Bad Rabbit ransomver is known for attacking its victims especially through the hacked Russian media pages. It spreads spontaneously on Internet and requires a ransom within the next 40 hours.
- **Fantom** - According to Tucu [18], the Fantom is considered as a new ransomware that looks like OS Windows update. After encryption, when to each encrypted file will be added the extension ".fantom", an Internet window will appear with other necessary steps.

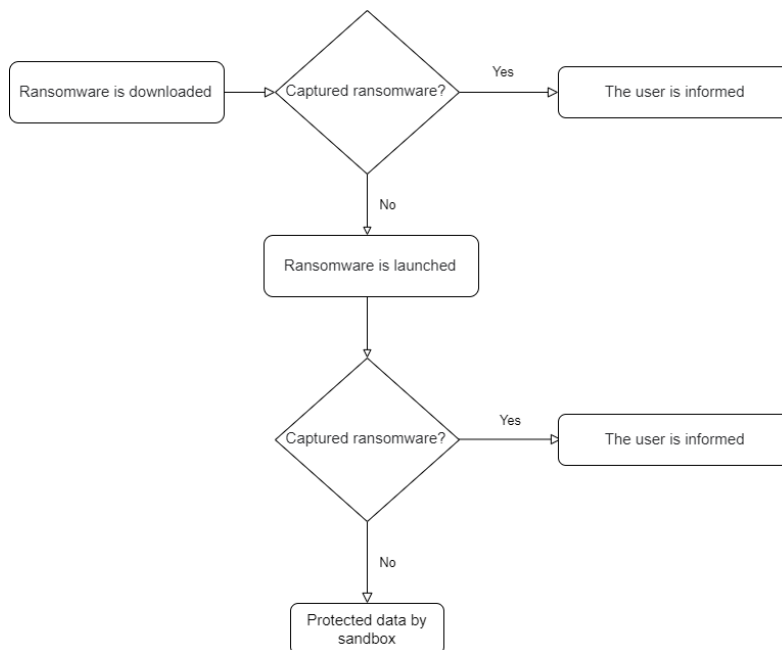


Fig. 4 Diagram of launched ransomware.

V. RESULTS

Based on the findings from the comparison of three available sandbox tools shown in table (Tab.1) is that as the best tool came out Sandboxie-Plus. It is a tool that does not require a lot of memory to operate and is suitable also for bulk samples testing. In addition, it can be used on Windows 7 and its later versions, making it more accessible to a larger number of users. From the table Tab.1 it is clear that Shade took second place in detecting harmful samples and revealed all samples during their execution. The worst was Windows Sandbox, which was able to detect all tested samples, however allowed the launch of four of them. The best out of the test was Sandboxie-Plus, which revealed four of the five ransomware test samples during their execution, and did not even allow the launch of one, as found at the beginning of testing.

As a result, we recommend Sandboxie-Plus to users without above-standard computer knowledge of using.

Table 1
Measurement values for individual sandboxes

Values obtained from measurements			
	Used sandbox tool		
Measurement	Sandboxie-Plus	Shade Sandbox	Windows Sandbox
RAM	17 MB	12 MB	457 MB
CPU	3,05 GHz	0,5 GHz	3,5 GHz
Browser	launch the default browser in the pc	must be added to the sandbox	built - in
Forced launch	folders and files	files after adding to the environment	files in the environment
Number of offered sandboxes	possibility to create more sandboxes	1	1
Number of possible running samples in one sandbox	multiple samples	multiple samples	multiple samples
Sample WannaCry	revealed	revealed	running in the environment
Sample Jigsaw	interrupted launch right at the beginning	revealed	revealed
Sample Petya	revealed	revealed	running in the environment
Sample Bad Rabbit	revealed	revealed	running in the environment
Sample Fantom	revealed	revealed	running in the environment

VI. CONCLUSION

During the creation of the work we gained new theoretical knowledge in the field of code analysis. We learned basic information about different types of analysis, such as static or hybrid analysis. But mainly, we've focused on a dynamic type of analysis that analyzes the code as it runs and looks for vulnerabilities that it might point to.

The final result obtained from our work was the finding that the Sandboxie-Plus tool performed best in our comparison. It does not require too much memory to operate. As for the launch effort several samples at once, it can create multiple sandbox instances and run multiple samples at the same time. This made it a suitable tool for bulk sample testing. Due to its ability to install in Windows 7 and its newer versions, it is considered more accessible to users. In case of ransoms launching, four samples were detected during their execution and one, a Jigsaw sample, even at the stage of its launch. Thanks to that it is a honoured data protection tool, which can prevent us from the loss of data, finance and time.

Based on the actually obtained results of the work, we can say that testing and comparing dynamic code analysis tools makes sense. With development of tools also malicious files are being developed at the same time. Therefore, it is necessary to give the tools to test at regular intervals and look for the best ones.

ACKNOWLEDGEMENT

This publication has been published with the support of the Operational Program Integrated Infrastructure within project: Research in the SANET Network and Possibilities of Its Further Use and Development (ITMS code: 313011W988), co-financed by the ERDF.

REFERENCES

- [1] PVS-Studio, "Static code analysis, [online]," 2015. [Online]. Available: <https://pvs-studio.com/en/blog/terms/0046/>
- [2] B. Chess and J. West, *Secure Programming with Static Analysis*, 1st ed. Addison-Wesley Professional, 2007.
- [3] ESEC/FSE-7: Proceedings of the 7th European Software Engineering Conference Held Jointly with the 7th ACM

- SIGSOFT International Symposium on Foundations of Software Engineering. Berlin, Heidelberg: Springer-Verlag, 1999.
- [4] T. Mezquita, "Dynamic code analysis, [online]," 2020. [Online]. Available:<https://cyberhoot.com/cybrary/dynamic-code-analysis/>
 - [5] D. Formánek, "Nástroje pro statickou a dynamickou analýzu kódu se zaměřením na bezpečnostní chyby, [online]," p. 72, 2014. [Online]. Available:<https://www.esecurityplanet.com/endpoint/sandboxing-advancedmalware-analysis/>
 - [6] S. Ingalls, "Sandboxing: Advanced malware analysis in 2022, [online]," 2021. [Online]. Available:<https://is.muni.cz/th/pngje/bp.pdf>
 - [7] K. A. Roundy and B. P. Miller, *Hybrid Analysis and Control of Malware*. Springer, Berlin, Heidelberg, 2010.
 - [8] L. Židzik, "Algoritmy detekcie malvéru na báze strojového učenia a statickej analýzy kód, [online]," p. 71, 2019. [Online]. Available:https://www.itspy.cz/wp-content/uploads/2019/10/29_prace.pdf
 - [9] S. Jamalpur, Y. S. Navya, and P. Rája, *Dynamic Malware analysis Using Cuckoo Sandbox*, 2018.
 - [10] L.Kaspersky, "Technology sandbox malware detection workflow,[online],"2022.[Online]. Available: <https://www.kaspersky.com/enterprisesecurity/wikisection/products/sandbox?fbclid=IwAR3mBu6q0KQC6YDxrsvsKiQFdcuHAWBgaGjUGXV6fggLqulyM9s2TaLOrw>
 - [11] G. Phillips, "The best sandbox tools to safely test windows programs, [online]," 2018. [Online]. Available: <https://www.makeuseof.com/tag/windows-sandbox-tools/>
 - [12] J. Hillier, "Shade sandbox: New sandboxing app, [online]," 2016. [Online]. Available:<https://davescomputertips.com/shade-sandbox-new-sandboxing-app/>
 - [13] Microsoft, "Windows sandbox, [online]," 2022. [Online]. Available:<https://docs.microsoft.com/en-us/windows/security/threatprotection/windows-sandbox/windows-sandbox-overview>
 - [14] M. Akbanov and V. G. Vassilakis, *WannaCry Ransomware: Analysis of Infection, Persistence, Recovery Prevention and Propagation Mechanisms.*, 2019.
 - [15] S. Cooper, "What is jigsaw ransomware how to protect against it? [online]," 2021. [Online]. Available: <https://www.comparitech.com/net-admin/jigsaw-ransomware/>
 - [16] J. S. Aidan, H. K. Verma, and L. K. Awasthi, *Comprehensive Survey on Petya Ransomware Attack*, 2017.
 - [17] A. Perekalin, "Bad rabbit: A new ransomware epidemic is on the rise, [online]," 2017. [Online]. Available: <https://www.kaspersky.com/blog/badrabbit-ransomware/19887/>
 - [18] T. M. I. SERVICES., "Fake windows update is actually fantom ransomware encrypting your files, [online]," 2022. [Online]. Available:<https://tucu.ca/fake-windows-update-fantom-ransomware/>

Využitie SPICE modelov v prostredí Matlab/Simulink

¹Marek PÁSTOR, ²Makar GOLOVKIN

^{1,2}Katedra elektrotechniky a mechatroniky, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Košice, Slovenská republika

¹marek.pastor@tuke.sk, ²makar.golovkin@student.tuke.sk

Abstrakt — Článok sa zaoberá možnosťou využitia SPICE modelov výkonových polovodičových súčiastok v prostredí Matlab/Simulink. Porovnáva rôzne možnosti simulácie výkonových tranzistorov prostredníctvom knižnice Simscape s ohľadom na komplexnosť modelu a simulované spínacie vlastnosti tranzistorov.

Kľúčové slová — model, Simulink, SPICE, tranzistor

I. ÚVOD

Počítačová simulácia a využívanie počítačových modelov pri návrhu elektronických systémov je dnes štandardom [1]. V oblasti výkonovej elektroniky je k dispozícii široká ponuka programov pre simulovanie vlastností výkonových polovodičových meničov. Štandardom v tejto oblasti je program OrCAD/PSpice (Personal Simulation Program with Integrated Circuit Emphasis) od firmy Cadence. Výrobcovia polovodičových súčiastok poskytujú počítačové modely vo formáte SPICE, čím umožňujú jednoduchú implementáciu modelu v simulácii. SPICE modely sú natoľko rozšírené a štandardizované, že ich využívajú aj iné simulačné programy. Nevýhodou programu OrCAD/PSpice je, že neumožňuje jednoduchú simuláciu riadiacich štruktúr meničov pre lineárne alebo nelineárne riadenie. Moderné riadiace metódy, ako sú fuzzy riadenie, prediktívne riadenie, sklzové riadenie a pod., tak nie je možné pri simulácii meniča aplikovať. Moderné riadiace algoritmy, vyžadujúce použitie mikrokontrolérov resp. digitálnych signálových procesorov alebo FPGA (Field Programable Gate Array) umožňujú výrazne zlepšiť vlastnosti výkonových polovodičových meničov [2]. Z toho vyplýva nutnosť simulácie meniča ako celku skladajúceho sa nielen z výkonovej topológie, ale aj z riadenia.

II. VYUŽITIE SPICE MODELOV V MATLABE

A. Simulink

Program Matlab (MATrix LABoratory) umožňuje jednoduchú realizáciu maticových výpočtov, prácu s dátami, implementovanie algoritmov, tvorbu užívateľských rozhraní a interakciu s inými programami. Obsahuje Simulink, čo je grafické rozhranie pre modelovanie, simulovanie a analýzu viacdoménových dynamických systémov. Spolu s knižnicou toolboxov umožňuje jednoduché implementovanie pokročilých riadiacich štruktúr. Simulink je tak vo veľkom využívaný v oblasti automatizácie, riadenia a digitálneho spracovania signálov pre viacdoménové simulácie a návrh na základe modelu. Simulink zároveň umožňuje priame generovanie programu v jazyku C. Je tak vhodný pre návrh lineárnych aj moderných nelineárnych riadiacich štruktúr aj pre oblasť výkonovej elektroniky.

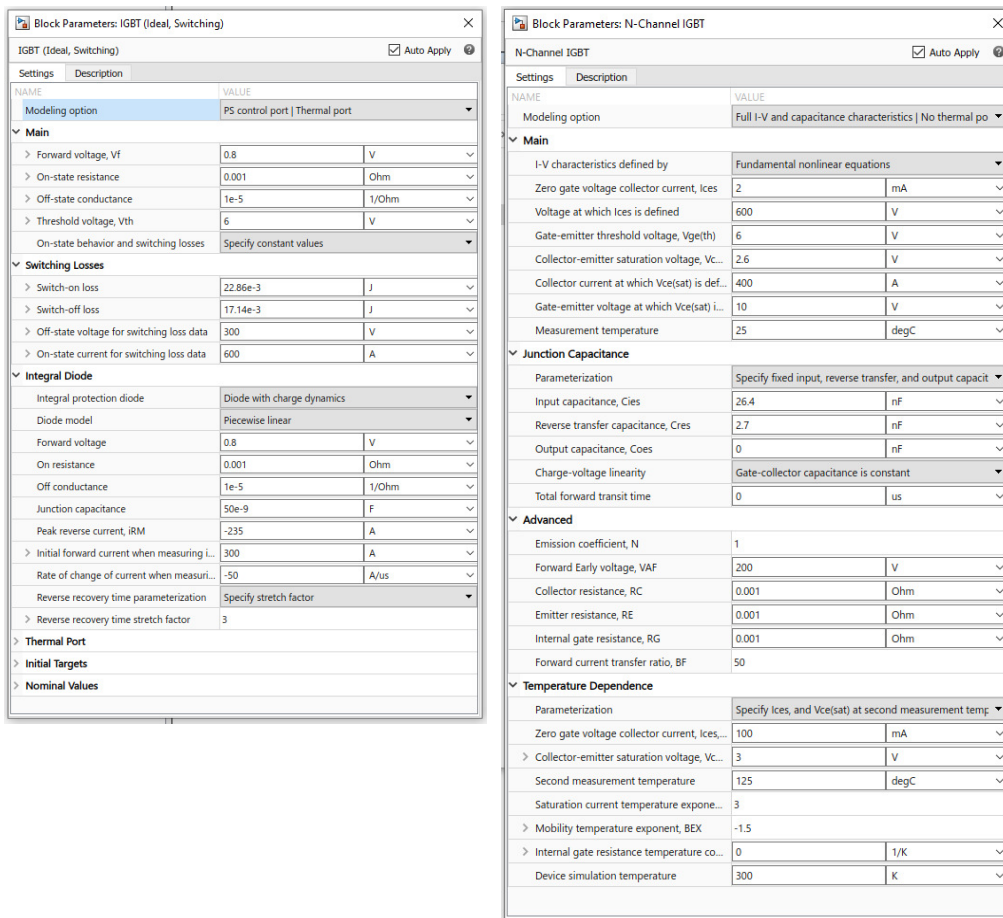
B. Toolbox Simscape Electrical

Samotný Simulink neumožňuje simulovanie výkonových polovodičových súčiastok. Umožňuje to však toolbox Simscape v rozšírení *Simscape Electrical*, ktorý obsahuje základné modely polovodičových spínačov ako sú tyristory, diódy a tranzistory ako aj hotových zapojení súčiastok v podobe základných výkonových polovodičových meničov [3]. V spojení z knižnicami určenými pre návrh riadiacich štruktúr a samotným Matlabom tak vznikol silný nástroj pre návrh riadenia výkonových meničov. Úroveň simulácie výkonového meniča však môže byť rôzna. Toolbox Simscape podporuje tri úrovne simulácie výkonových polovodičových súčiastok. Prvou je simulácia na úrovni ideálneho spínača, druhou je simulácia so zohľadnením

dynamických vlastností a tretou je simulácia na úrovni SPICE modelu. Všetky tri sú ďalej uvedené na príklade výkonového tranzistora typu IGBT.

C. Model ideálneho tranzistora IGBT

Základný model tranzistora IGBT v rámci knižnice Simscape Electrical, označený *IGBT Ideal, Switching*, (Obr. 1 a)) zohľadňuje základné statické vlastnosti ako úbytok napätia v priepustnom smere V_f a spínacie vlastnosti ako prahové napätie V_{th} . Úbytok napätia v priepustnom smere je buď konštantný alebo môže byť funkciou prúdu a teploty. V prípade antiparalelnej diódy umožňuje model simulovanie aj jej dynamických vlastností a spätného zotavenia. Spínacie straty sú simulované ako energia pri zapnutí a vypnutí v závislosti od kolektorového napätia a prúdu. Tieto hodnoty je možné získať z katalógu tranzistora. Nakoľko pri zapnutí tranzistora nie je známy jeho konečný kolektorový prúd a pri vypnutí nie je známe jeho konečné kolektorové napätie, tak model vypočítava straty tranzistora s oneskorením jedného spínacieho cyklu.



a) Základný model IGBT tranzistora v prostredí Simscape

b) Rozšírený model IGBT tranzistora v prostredí Simscape

Obr. 1 Parametre ideálneho modelu tranzistora *IGBT Ideal, Switching* a *N-Channel IGBT*

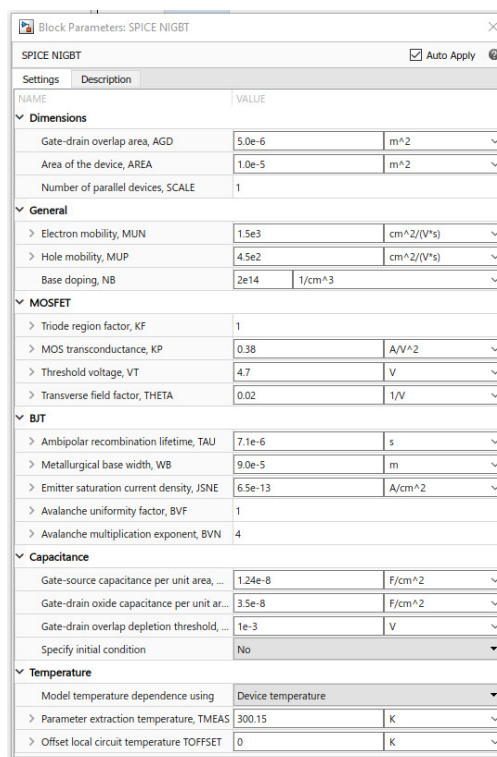
D. Dynamický model tranzistora IGBT

Dynamický model tranzistora IGBT je v knižnici Simscape Electrical označený *N-Channel IGBT* (Obr. 1 b)). Tento model má dve možnosti. Prvou je detailná simulácia volt-ampérových charakteristík tranzistora so zohľadnením vstupných a výstupných kapacít tranzistora. Pri tejto možnosti je zohľadnená aj teplotná závislosť parametrov. Model je vhodný pre simulovanie spínacích vlastností a predikovanie spínacích strát. Obsahuje model tranzistora MOSFET a IGBT podobne, ako je tomu v prípade náhradnej schémy IGBT tranzistora alebo využíva 2-D alebo 3-D tabuľku parametrov (look-up table). Vstupná kapacita C_{GE} je simulovaná ako konštantná a kapacita C_{GC} je simulovaná ako konštantná alebo nelineárne závislá od kolektorového napätia. To umožňuje simulovať Millerove platá. Model však nesimuluje teplotnú závislosť kapacít tranzistora.

Druhou možnosťou je simulovanie zjednodušených volt-ampérových charakteristík, podobne ako je tomu v prípade základného modelu *IGBT Ideal Switching*. Simulácia spínania tranzistora je založená na modely riadenom udalosťami a tranzistor sa vždy nachádza v jednom zo štyroch stavov: zapnutý, vypnutý, zapínajúci alebo vypínajúci. Táto možnosť simulácie je vhodná v prípade dôrazu na rýchlosť simulácie s postačujúcimi základnými dynamickými a statickými vlastnosťami tranzistora. Prúdový chvost tranzistora nie je simulovaný. Podobe ako v prípade základného modelu *IGBT Ideal Switching* sú straty tranzistora aj v tomto prípade počítané s oneskorením jedného spínacieho cyklu.

E. SPICE model tranzistora IGBT

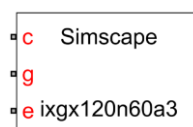
SPICE model poskytuje najdetailnejší model tranzistora využívaný špecializovanými simulačnými programami pre elektronické simulácie. SPICE model je štandardizovaný model. Simulink umožňuje v toolboxe *Simscape Electrical* pracovať so SPICE modelmi dvoma spôsobmi. Prvým je využiť blok umožňujúci zadávanie štandardizovaných parametrov SPICE modelu (Obr. 2). Pre využitie tohto modelu je však potrebné mať k dispozícii parametre, ktoré nie sú dostupné v katalógovom liste tranzistora.



Obr. 2 Parametre SPICE modelu tranzistora IGBT *SPICE NIGBT*

Parametre je možné získať napr. pomocou programu OrCAD Model Editor na základe údajov z katalógového listu tranzistora. SPICE model tranzistora IGBT obsahuje 17 parametrov, ktoré sú zobrazené na Obr. 2 [4].

Výrobcovia polovodičových súčiastok často poskytujú hotové SPICE modely pre uľahčenie práce, často vo formáte priamo vhodnom pre OrCAD PSpice. Takýto model môže byť vo forme SPICE obvodu označovaného subcircuit. *Simscape Electrical* umožňuje konverziu SPICE obvodu do bloku použiteľného pri simulácii v prostredí Simulink. Na konverziu slúži príkaz *subcircuit2ssc*, ktorého vstupom je SPICE netlist a výstupom je súbor pre *Simscape* (Obr. 3). Takýto postup je jednoduchší, ako vyhľadávanie, prípadne generovanie, všetkých parametrov pre definovanie SPICE modelu tranzistora IGBT.



Obr. 3 Simscape model IGBT tranzistora vytvorený na základe SPICE modelu pomocou príkazu *subcircuit2ssc*

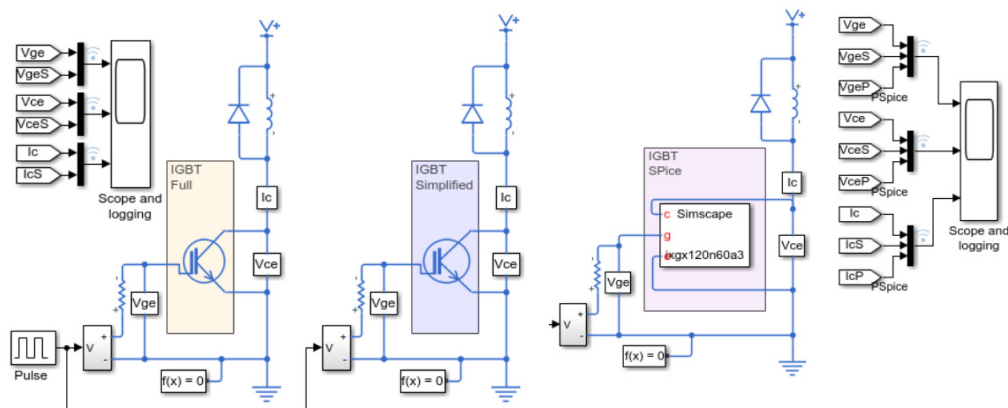
III. SIMULÁCIA IGBT TRANZISTORA

Pre vzájomne porovnanie dynamických vlastností troch modelov tranzistora IGBT v prostredí toolboxu Simscape Electrical bol vytvorený simulačný model podľa vyššie opísaných možností. Pre účely simulácie bol zvolený tranzistor IXGX120N60A3 od firmy IXYS. Jeho základné parametre sú uvedené v Tabuľke 1.

Tabuľka 1
Základné parametre tranzistora IXGX120N60A3 [5]

Parameter	Hodnota
Kolektorové napätie U_{CE}	600 V
Kolektorový prúd pre zohriaty tranzistor I_{C110}	120 A
Saturačné napätie $U_{CE(sat)}$	$\leq 1,35$ V
Maximálna výkonová strata P_D	780 W
Náboj hradla Q_G	450 nC

Tri vytvorené modely tranzistora sú znázornené na Obr. 4. Na obrázku Obr. 4 a) sú znázornené dva modely obsahujúce bloky z knižnice Simscape doplnené o hodnoty z katalógového listu tranzistora. Blok IGBT Simplified obsahuje základný model IGT tranzistora. Blok IGBT Full obsahuje rozšírený model IGBT tranzistora. Premenné v blokoch boli nastavené na základe údajov z katalógového listu tranzistora. V prípade základného modelu IGBT tranzistora to bol vektor kolektorového prúdu a zodpovedajúcich úbytkov napätia, prahové napätie, oneskorenie pri zapnutí, doba nárastu, oneskorenie pri vypnutí a doba poklesu. Pre rozšírený model boli statické vlastnosti modelované základnými rovnicami modelu tranzistora. Do modelu boli pridané hodnoty kapacít tranzistora. Model na Obr. 4 b) je tvorený blokom, ktorý bol vytvorený zo SPICE súboru od výrobcu tranzistora použitím príkazu *subcircuit2ssc*.



a) Základný a rozšírený model IGBT tranzistora v prostredí Simscape

b) SPICE model IGBT tranzistora v prostredí Simscape

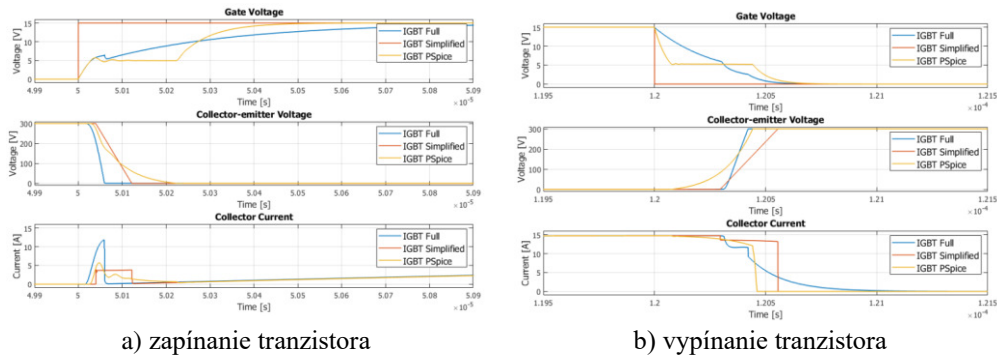
Obr. 4 Programové schémy pre simuláciu dynamických vlastností tranzistora IXGX120N60A3 v Simulinku [6]

A. Simulované dynamické vlastnosti tranzistora

Pre porovnanie jednotlivých modelov IGBT tranzistora boli modely zapojené v obvode impulzového znižovacieho meniča s parametrami uvedenými v Tabuľke 2.

Tabuľka 2
Parametre znižovacieho impulzového meniča

Parameter	Hodnota
Napájacie napätie U_1	300 V
Max. prúd záťaže I_{OUT}	15 A
Spinacia frekvencia f_{sw}	20 kHz

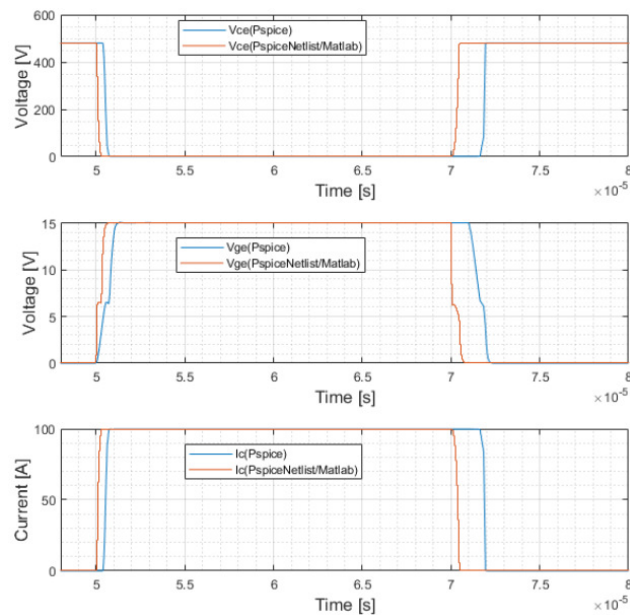


Obr. 5 Časové priebehy tranzistora pri spínaní indukčnej záťaže (hore: hradlové napätie, v strede: kolektorové napätie, dole: kolektorový prúd) [6]

Model nulovej diódy v znižovacom impulzovom meniči zohľadňoval spätné zotavenie diódy. Výsledky simulácie zapnutia a vypnutia IGBT tranzistora sú uvedené na Obr. 5. Základný model tranzistora nezohľadňuje proces nabíjania vstupnej kapacity tranzistora, pri zapínaní a vypínaní tranzistora uvažuje s prechodovým dejom definovaným pomocou zadaných spínacích časov tranzistora a nezohľadňuje nadprúd od spätného zotavenia nulovej diódy. Rozšírený model IGBT tranzistora už vďaka uvažovaniu kapacít simuluje aj ich nabíjanie a vybíjanie, avšak len v zjednodušenej forme. Pre lepšie priblíženie k reálnym priebehom tranzistora by bolo potrebné doladiť hodnoty kapacít. Model tranzistora získaný zo SPICE modelu verne simuluje priebeh hradlového napätia tranzistora, priebeh kolektorového napätia ako aj kolektorového prúdu. Vďaka tomu dokáže najhodnovernejšie simulovať spínacie straty tranzistora. Z výsledkov simulácie (Obr. 5) je vidieť, že základný model poskytuje len veľmi zjednodušené výsledky a nie je vhodný v prípadoch keď je potrebné simulovať spínacie procesy tranzistora.

B. Porovnanie s OrCAD/PSpice

Pre verifikáciu modelu tranzistora vytvoreného zo SPICE modelu pomocou príkazu *subcircuit2ssc* pre toolbox Simscape bol originálny SPICE model IGBT tranzistora simulovaný v programe OrCAD PSpice. Výsledky simulácie boli porovnané s blokom vytvoreným pre Simscape. Tranzistor bol zapojený v rovnakom impulzovom znižovacom meniči. Výsledky simulácie sú znázornené na Obr. 6. Ako je vidieť, tak oba modely simulujú dynamický proces zapínania a vypínania tranzistora veľmi podobne. Pozorovateľný rozdiel je len v časoch oneskorenia pri zapnutí a vypnutí. Model tranzistora vytvorený pomocou príkazu *subcircuit2ssc* pre toolbox Simscape je tak možné používať pre presné simulácie so zohľadnením spínania výkonového tranzistora. Je však potrebné zdôrazniť, že výsledky simulácie uvedené na Obr. 5 boli získané z dvoch odlišných programov.



Obr. 6 Časové priebehy tranzistora pri porovnaní SPICE modelu o programe orCAD a Simscape modelu odvodeného zo SPICE modelu (hore: kolektorové napätie, v strede: hradlové napätie, dole: kolektorový prúd) [6]

IV. ZÁVER

V článku sú opísané a porovnané možnosti simulácie výkonových polovodičových súčiastok v prostredí Matlab/Simulink s využitím toolboxu *Simscape Electrical*. Tento toolbox umožňuje simulovať výkonovú časť meniča a s využitím ostatných toolboxov implementovať pokročilé riadiace algoritmy a spracovanie dát, ktoré nie je možné realizovať v špecializovaných simulačných programoch určených pre simuláciu elektronických obvodov, napr. OrCAD/PSpice. Toolbox *Simscape Electrical* umožňuje simulovať výkonové polovodičové súčiastky v troch úrovniach zložitosti. To umožňuje zvoliť si medzi presnosťou simulácie na úrovni spínacích časov a rýchlosťou simulácie. Z porovnania základného modelu, rozšíreného modelu a modelu generovaného z SPICE modelu IGBT tranzistora vyplýva, že pre základné simulácie dynamických dejov pri spínaní tranzistora je postačujúce používať správne natavený rozšírený model tranzistora. Pre detailné simulácie spínacích dejov je vhodné použiť model generovaný na základe SPICE modelu od výrobcu tranzistora. Takýto model je svojimi vlastnosťami porovnateľný s modelom implementovaným v prostredí OrCAD/PSpice.

REFERENCIE

- [1] G.T. Oziemkiewicz, "Implementation and Development of the NIST IGBT Model in a SPICE-based Commercial Circuit Simulator," Engineer's Thesis, University of Florida, December 1995.
- [2] M. Leso, J. Zilkova, P. Girovsky, "Development of Simple Fuzzy Logic Controller for DC-DC Converter", *Proc. IEEE 18th International Power Electronics and Motion Control Conference (PEMC)*, 2018, pp. 86–93
- [3] Simscape Electrical – User Guide, Mathworks, verzia 7.7, 2022. Dostupné online: https://www.mathworks.com/help/pdf_doc/physmod/sps/sps_ug.pdf
- [4] PSpice Reference Guide, Cadence Design Systems, druhé vydanie, 2000. Dostupné online: https://www.seas.upenn.edu/~jan/spice/PSpice_ReferenceguideOrCAD.pdf
- [5] Katalógový list tranzistora IXGX120N60A3, IXYS, verzia 02/2009. Dostupné online: <https://www.littelfuse.com/media?resourcetype=datasheets&itemid=c6af7d69-5684-4694-865e-d9b7a8f5c0c5&filename=littelfuse-discrete-igbts-pt-ixg-120n60a3-datasheet>
- [6] M. Golovkin, „Simulácia výkonových meničov v Matlabe“, bakalárska práca, TUKE, 2021

Návrh Ultra-DWDM systému s optickým zosilňovačom SOA

¹ Norbert ZDRAVECKÝ, ² Luboš OVSEŇÍK, ³ Maroš LAPČÁK, ⁴ Jakub ORAVEC, ⁵ Samuel ANDREJČÍK

^{1,2,3,4,5} Katedra elektroniky a multimediálnych telekomunikácií, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

¹ norbert.zdravecky@tuke.sk, ² lubos.ovsenik@tuke.sk, ³ maros.lapcak@tuke.sk, ⁴ jakub.oravec@tuke.sk, ⁵ samuel.andrejcik@tuke.sk

Abstrakt — Tento príspevok sa zaoberá návrhom a analýzou 32-kanálového Ultra-DWDM systému s použitím polovodičového optického zosilňovača (SOA). Návrh skúma zosilnenie jednotlivých kanálov pri použití SOA v Ultra-DWDM. Pozoruje nežiadúce vplyvy lineárnych a nelineárnych javov, ktoré sú hlavným faktorom ovplyvňujúcim kapacitu a prenosovú vzdialenosť optického komunikačného systému.

Kľúčové slová — BER, DWDM, nelineárne javy, SOA, ultra-DWDM

I. ÚVOD

Rastúci dopyt po vysokorychlostných a vysokokapacitných systémoch na dlhé vzdialenosti prinútil optickú komunikačnú technológiu zvažovať riešenie systémov multiplexovania s hustým vlnovým delením DWDM (Dense Wavelength Division Multiplexing). DWDM umožňuje kombináciu množstva nezávislých vlnových dĺžok prenášajúcich informácie na rovnakom vlákne vďaka širokej spektrálnej oblasti v ktorej je možné efektívne prenášať optické signály [1]. Pre maximalizáciu výkonu takýchto optických sietí sa zaradzujú do systémov optické zosilňovače. Všetky typy optických zosilňovačov majú svoje výhody aj nevýhody. Kritickou oblasťou výskumu DWDM sietí sa stali vhodné stratégie použitia a umiestnenia zosilňovačov [5]. Návrh komunikačného systému a optimalizácia parametrov súvisia s prenosom dát v rámci vlastností kanála. Kapacita optického kanála je zvyčajne ovplyvnená tvarom signálu, disperznými a nelineárnymi charakteristikami optického vlákna [3]. Pre dosiahnutie najlepších možných parametrov bez degradácie signálu je dôležité pochopenie, modelovanie a charakterizovanie správania viackanálových optických sietí.

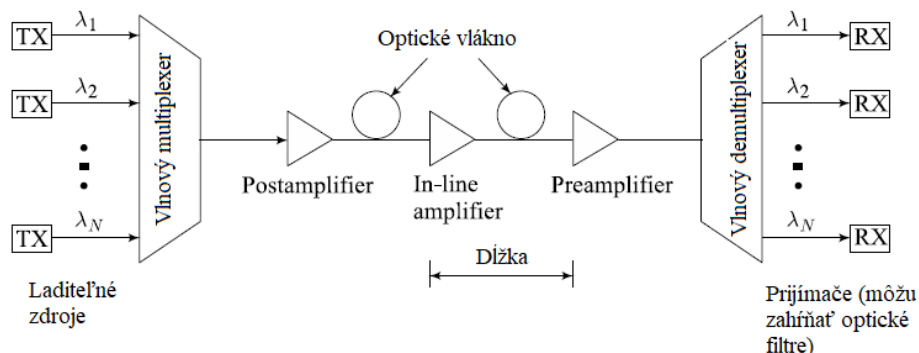
Technológia DWDM používa multiplexor (transpondéry) vo vysielacom zariadení pre zloženie signálov a demultiplexor v prijímači pre ich rozdelenie.

II. OPTICKÉ VLÁKNOVÉ PRENOSOVÉ SYSTÉMY S HUSTÝM VLNOVÝM DELENÍM

Optické systémy s hustým vlnovým delením sú možným riešením pre viackanálové optické vláknové komunikácie. Odporúčania pre DWDM systémy sú zhrnuté v štandarde ITU-T G.694.1. Referenčná vlnová dĺžka pre DWDM kanály je 1552,5 nm (193,10 THz). Komerčne DWDM systémy dokážu multiplexovať až 128 optických nosných vln [2].

Použitie kanálového rozstupu 50 GHz pri bitovej rýchlosti 40 Gb/s vo frekvenčnom rozsahu 186-201 THz zvýšilo spektrálnu účinnosť systémov IM/DD na 0,8 b/s/Hz. Táto implementácia vyžaduje použitie stabilných a vysoko kvalitných laserových svetelných zdrojov s regulovanou teplotou a vlnovou dĺžkou [3].

Nižšie kanálové rozstupy ako 50 GHz stále nie sú v chrbticových optických sieťach veľmi rozšírené. V prípade systémov DWDM môžu byť nelineárne javy dôležité aj pri miernych výkonoch a bitových rýchlostiach. Existujú dve kategórie nelineárnych javov [4]. Prvá kategória je spôsobená závislosťou indexu lomu na intenzite šíriaceho sa signálu vláknom. Najdôležitejšie nelineárne javy v tejto kategórii sú vlastná fázová modulácia (SPM), krížová fázová modulácia (XPM) a štvorvlnové zmiešavanie (FWM) [6,7]. Optická sieť DWDM je znázornená na Obr. 1.



Obr. 1 Schéma DWDM systému s rôznymi možnosťami zapojenia zosilňovača.

Na strane prijímača sa nachádza laserový zdroj generujúci vlnu, na ktoré sa moduluje elektrický signál do optického dátového toku. Signály z jednotlivých laserov sa následne multiplexujú a prenáša optickou trasou tvorenou jednovidovým optickým vláknom (SMF – Single Mode Fiber) a optickým zosilňovačom. Optický výkonový zosilňovač (v našom prípade SOA) neguje vložné straty multiplexora na strane vysielača [4,10].

Optický predzosilňovač zosilňuje optický signál tak, aby bol detekovateľný prijímačom. Pri prenose optickým vláknom na dlhé vzdialenosti je zvyčajne použitý aj linkový zosilňovač. Hlavnou úlohou je kompenzovať straty spôsobené prenosom cez optickú trasu. Na prijímacom konci je potrebný demultiplexor vlnových dĺžok na oddelenie optických signálov do vhodných detekčných kanálov s rôznymi vlnovými dĺžkami na spracovanie signálu [8,9].

Keď sa N kanálov s bitovými rýchlosťami B_1, B_2, \dots, B_N prenáša súčasne cez vlákno dĺžky L , tak súčin celkovej bitovej rýchlosti a vzdialenosti B_L je:

$$B_L = (B_1 + B_2 + \dots + B_N)L \quad (1)$$

III. NELINEÁRNE JAVY V DWDM SYSTÉMOCH

Vznik nelineárnych javov je podmienený veľkým svetelným výkonom vo vlákne. Problém je v tom, že vlákna majú veľmi malý prierez jadra a prichodom vlnového multiplexu sa do dlhších trás začali začleňovať optické zosilňovače, ktoré niekoľkonásobne zvyšujú výkon vo vlákne. Pri prenosových rýchlostiach nad 10 Gb/s narážame na obmedzenia spojené s nelineárnymi javmi. Vznikajú v dôsledku interakcie svetelných vln s fonónmi v optických vláknoch. Nelineárne rozptylové javy sú stimulovaný Ramanov rozptyl (SRS) a stimulovaný Brillouinov rozptyl (SBS).

SRS je hlavným problémom v systémoch multiplexovania s vlnovou dĺžkou, pretože môže vyvolať presluchy medzi rôznymi kanálmi. V takom prípade kanály s vyššou nosnou frekvenciou dodávajú časť svojho výkonu kanálom s nižšou frekvenciou [6]. Rovnaký efekt však môže nájsť užitočné aplikácie, konkrétne na efektívne zosilnenie vysielačských signálov a na generovanie nových frekvencií.

Kerrov efekt vzniká v dôsledku závislosti indexu lomu optického vlákna na intenzite svetelného lúča [3]. Index lomu jadra optického vlákna je možné vyjadriť ako:

$$n = n_0 + n_2 \cdot \frac{P}{A_e}, \quad (1)$$

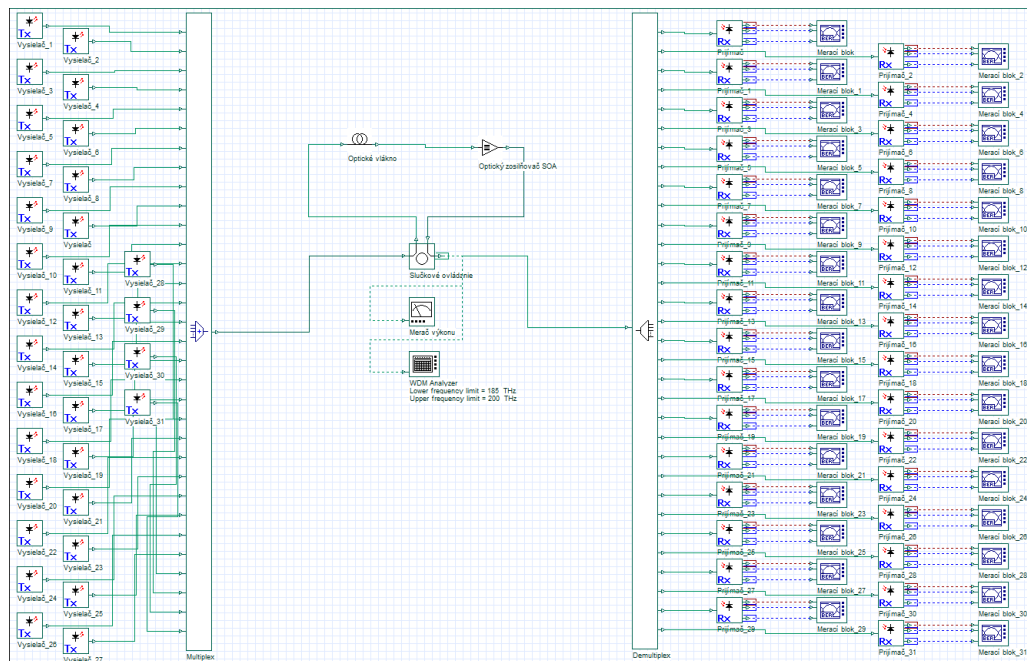
kde n_0 je index lomu jadra optického vlákna, n_2 , P je výkon naviazaný na optické vlákno a A_e je efektívna plocha jadra optického vlákna. Nelineárne javy na základe Kerrovho efektu sú krížová fázová modulácia XPM (Cross Phase Modulation), vlastná fázová modulácia SPM (Self Phase Modulation), štvorvlnové zmiešavanie FWM (Four Wave Mixing).

Hlavným parametrom je nelineárny koeficient γ , ktorý udáva veľkosť nelineárneho účinku pre SPM. Tento parameter je daný:

$$\gamma = \frac{2\pi n_2}{\lambda A_e} \quad (2)$$

IV. NÁVRH ULTRA-DWDM SYSTÉMU S OPTICKÝM ZOSILŇOVAČOM SOA

Dizajn nášho Ultra-DWDM systému vychádza z dlhotrvajúceho výskumu, poznatkov a vedomostí. Kapacita je 0,32 Tb/s, čo znamená prenosovú rýchlosť 10 Gb/s na 1 kanál. Dĺžka prenosovej trasy tvorenej jednovidovým optickým vláknom je od 35 km do 210 km. Výkon lasera sa menil od -3 dBm po 3 dBm. Schéma navrhnutého systému je na Obr. 2. Ďalšie parametre tohto modelu sú uvedené v Tab. 1. Analyzoval sa výkon na výstupe pre tri hodnoty vstupného výkonu.



Obr. 2 Navrhnutý model Ultra-DWDM systému.

Optické vlákno s dĺžkou 35 km má Q faktor najvyšší pre hodnotu výkonu lasera -3 dBm. Rozdiel hodnôt Q faktoru pri výkonoch 0 dBm a 3 dBm je veľmi malý. Pri dĺžke optického vlákna do 105 km pri výkone 3 dBm je krivka lineárne klesajúca, čím sa výrazne odlišuje od ďalších vstupných výkonov. Pre všetky tri výkony je hraničná hodnota prenosovej vzdialenosti 70 km, kde krivka razantne klesá až na vzdialenosť 105 km.

Tabuľka 1 Parametre navrhnutého Ultra-DWDM systému

Vysielač – 32 kanálov	
Bitová rýchlosť	10 Gb/s
Počet vzoriek na 1 bit	64
Počiatková frekvencia	193,1 THz
Medzikanálové rozostupy	200 GHz
	-3 dBm
Vyžarovací výkon	0 dBm
	3 dBm
Prenosová trasa	
	35 km
Dĺžka SMF	70 km
	105 km
	210 km
Hodnota chromatickej disperzie SMF	17 ps/nm/km
Referenčná vlnová dĺžka	1550 nm
Útlm	0,2 dB/km
Disperzia - β_2	-20 ps ² /km
Polarizačná disperzia – PMD	0,2 ps/km
Efektívna oblasť vlákna A_e	80 μm^2
Nelineárny index lomu n_2	2,6e-020 m ² /W
Hodnota chromatickej disperzie DCF	-85 ps/nm/km

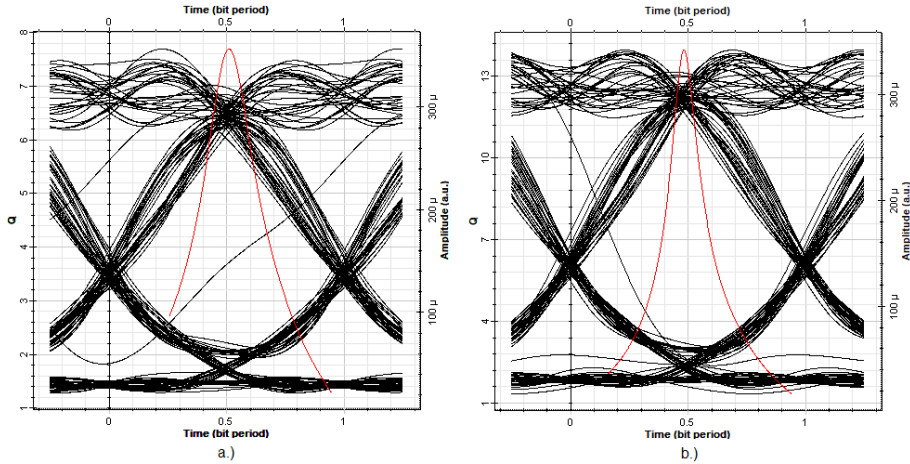
Po vykonaní simulácií pri rôznych vstupných výkonoch sme zaznamenali dosiahnuté hodnoty BER do grafickej závislosti pre -3 dBm na Obr. 3. Súčasný limit BER pre DWDM systémy je $1e-10$, čo po zlogaritmovaní predstavuje hodnotu -23,0259. To znamená, že DWDM systém je z hľadiska spoľahlivosti nevyhovujúci ak je hodnota BER prijatého signálu vyššia ako $1e-10$.

Tabuľka 2 Výsledky Ultra-DWDM pre výkon lasera -3 dBm.

DWDM kanál	BER	Q faktor
CW výkon -3 dBm		
5	$2,11 \times 10^{-26}$	10,5673
11	$1,70 \times 10^{-29}$	11,2162
15	$1,45 \times 10^{-32}$	11,8226
19	$9,60 \times 10^{-41}$	13,3125

21	$1,14 \times 10^{-37}$	12,7735
25	$2,27 \times 10^{-35}$	12,3547
29	$2,33 \times 10^{-41}$	13,4193

Hodnoty Q faktora pri výkone lasera -3 dBm a pri prenosovej dĺžke optického vlákna 35 km, 70 km, 105 km a 210 km je 14,31 dB, 12,68 dB a 0,2 dB. Tabuľka 2 zobrazuje hodnoty BER a Q faktora pre jednotlivé kanály, najlepšia hodnota Q faktora je pri 29 kanáli.

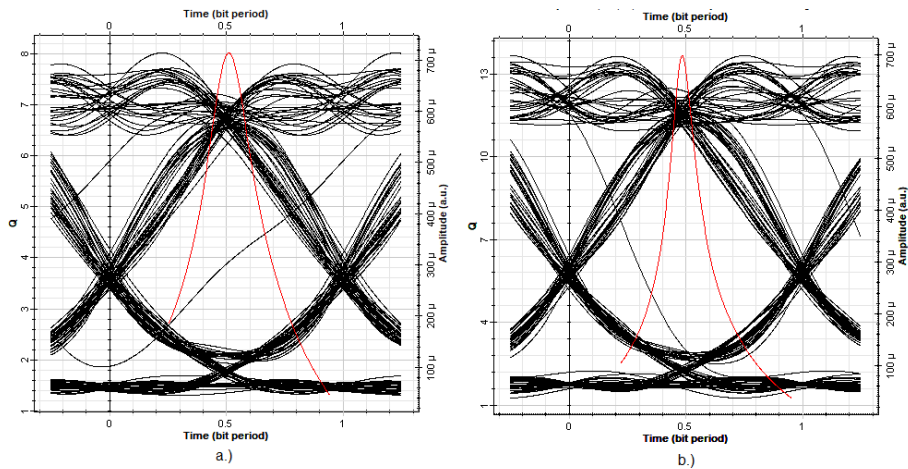


Obr. 3 Diagramy oka s využitím SOA zosilňovača pri výkone -3 dBm pre a.) 17. kanál, b.) 27. kanál

Tabuľka 3 Výsledky Ultra-DWDM pre výkon lasera 0 dBm

DWDM kanál	BER	Q faktor
CW výkon 0 dBm		
5	$8,32 \times 10^{-29}$	11,0743
11	$1,08 \times 10^{-27}$	10,8424
15	$4,36 \times 10^{-31}$	11,5345
19	$9,06 \times 10^{-43}$	13,6568
21	$1,87 \times 10^{-37}$	12,7347
25	$1,44 \times 10^{-36}$	12,5744
29	$1,07 \times 10^{-38}$	12,9574

Hodnoty Q faktora pri výkone lasera 0 dBm a pri prenosovej dĺžke optického vlákna 35 km, 70 km, 105 km a 210 km je 13,69 dB, 12,92 dB, 2,64 dB, 2,7 dB a 0,2 dB.

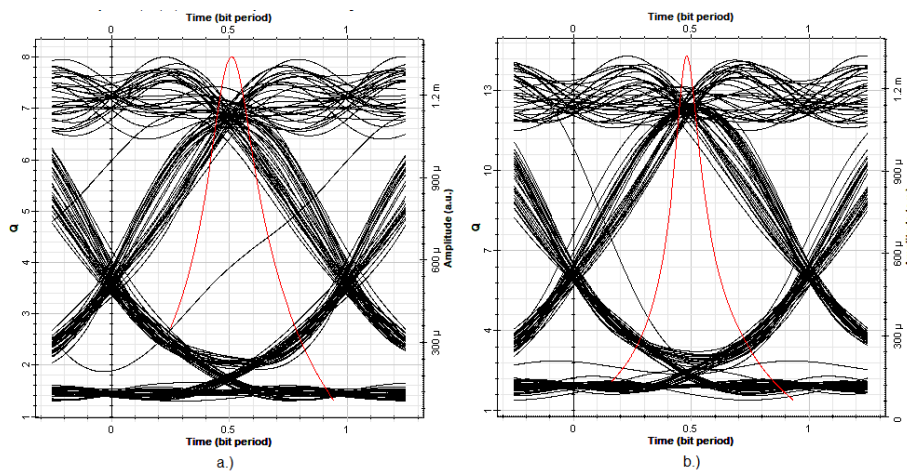


Obr. 4 Diagramy oka s využitím SOA zosilňovača pri výkone 0 dBm pre a.) 17. kanál, b.) 27. Kanál

Tabuľka 4 Výsledky Ultra-DWDM pre výkon lasera 3 dBm

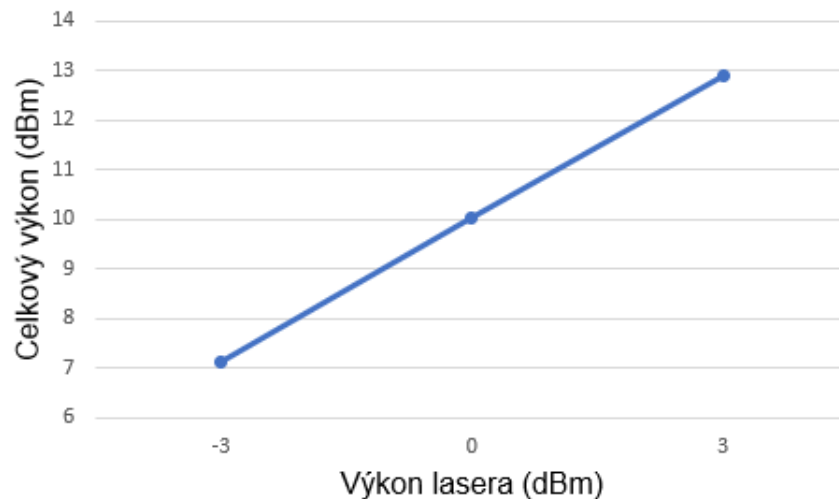
DWDM kanál	BER	Q faktor
CW výkon 3 dBm		
5	$5,55 \times 10^{-30}$	11,3147
11	$1,55 \times 10^{-36}$	12,5698
15	$2,06 \times 10^{-39}$	13,0813
19	$6,21 \times 10^{-44}$	13,8511
21	$1,70 \times 10^{-41}$	13,4424
25	$1,69 \times 10^{-36}$	12,5624
29	$2,43 \times 10^{-40}$	13,2443

Hodnoty Q faktora pri výkone lasera 3 dBm a pri prenosovej dĺžke optického vlákna 35 km, 70 km, 105 km a 210 km je 21,63 dB, 16,78 dB, 2,73 dB a 0,2 dB. Z hodnôt môžeme zistiť, že celkový výkon klesá s nárastom vzdialenosti vlákna, čo je spôsobené stratou signálu v optickom vlákne v dôsledku nelineárnych javov.



Obr. 5 Diagramy oka s využitím SOA zosilňovača pri výkone 3 dBm pre a.) 17. kanál, b.) 27. Kanál

Pri výkone lasera -3 dBm hodnoty chybovosti BER sa pohybovali v rozsahu $6,75 \times 10^{-15}$ až $1,14 \times 10^{-44}$. Pri narastajúcom výkone lasera na 3 dBm sa hodnoty BER pohybovali $5,77 \times 10^{-16}$ až $9,67 \times 10^{-47}$. Pri analýze systému sme pozorovali ako sa zmení celkový výkon na výstupe signálu pre tri rôzne hodnoty výkonu lasera pre SOA zosilňovač. Graf (Obr. 6) ukazuje, že hodnoty prijatého výkonu sú pri -3 dBm, 0 dBm a 3 dBm v tomto poradí 7,117 dBm, 10,043 dBm a 12,901 dBm. Môžeme vidieť, že krivka má narastajúci lineárny priebeh. Zmena však nastala pri narastajúcom výkone lasera, kde bola hodnota Q faktoru oveľa horšia pre rovnaký kanál.



Obr. 6 Závislosť celkového výkonu na výkon lasera pre SOA zosilňovač

V. ZÁVER

Počas simulácie experimentálneho návrhu Ultra-DWDM systému s využitím optického zosilňovača SOA v programovom prostredí OptiSystem bez použitia kompenzácie chromatickej disperzie sa zistilo, že hodnoty chybovosti BER a Q faktoru sú vysoko nadprahovými hodnotami. Pri meranej dĺžke optickej siete pri 105 km sa z grafu zistila vysoká medzisymbolová interferencia. Z týchto dôvodov sa simulovali experimentálne návrhy aj pri dĺžke optickej siete 35 km, 70 km a 210 km, kde slučkovým ovládaním v schéme sa vykonali tri opakovania. Výskum v tejto oblasti je nesmierne dôležitý a umožňuje dosahovať stále vyššie prenosové vzdialenosti DWDM systémov. Práca v budúcnosti pomôže pri nasadzovaní jednotlivých typov zosilňovačov v Ultra-DWDM sieťach.

Na základe dosiahnutých výsledkov možno konštatovať, že zosilňovače SOA je vhodné nasadzovať pri 10 Gb/s systémoch do vzdialeností 105 km.

POĎAKOVANIE

Táto práca bola podporená výskumným projektom FEI-2022-84 "Techniky spracovania dát vo vysokorychlostných prenosových systémoch".

LITERATÚRA

- [1] Gerd Keiser: Fiber Optic Communication. USA: Newton Center, MA, 2021, ISBN: 978-981-33-4665-9, 639 pp.
- [2] Devi Chadha: Optical WDM networks: From static to elastic networks. India: Indian Institute of Technology, 2019, ISBN: 9781119393269, 359 pp.
- [3] Agrawal, Govind P.: Fiber-Optic Communication Systems (Fifth Edition). New York: The Institute of Optics University of Rochester, 2021, ISBN: 9781119737377, 544 pp
- [4] A. Dochhan, et al., "Solutions for 80 km DWDM Systems", *Journal of Lightwave Technology*, 2015, pp. 491–499.
- [5] M. Singh, A review on hybrid optical amplifiers. *J. Opt. Commun.* 39(3), 267-272, 2018
- [6] G. P. Agrawal, "Nonlinear Fiber Optics", 2019, 631 pp.
- [7] B. A. Saleh, M. Teich, Fundamentals of Photonics (Third Edition). New York, 2019.
- [8] R. Ramaswami, et al., "Optical Networks", 2010, 928 pp.
- [9] K. Iizuka: Engineering Optics. Department of Electrical and Computer Engineering University of Toronto. Canada. 2019. ISBN: 978-3-319-69251-7, 795 pp.
- [10] S. Kheraliya, Ch. Kumar: Comparative Study of Various Optical Amplifiers for 32-Channel WDM System. *Journal of Optical Communications*. 2018.

Predicting Human Behaviour in Online Games Using Neural Networks

¹Stanislav Ivanchyk, ²Martina Szabóová

^{1,2}Department of Cybernetics and Artificial Intelligence, Faculty of Electrical Engineering and Informatics Technical University of Košice, Slovak Republic

¹stanislav.ivanchyk@student.tuke.sk ²martina.szaboova@tuke.sk

Abstract – This paper describes deep learning to predict player behavior in an online slot game. The work consists of an introduction that describes the reasons for this work (such as game addiction and game design optimization) and formulates the objectives. The analytical part describes the problem in terms of deep learning, namely the problem of predicting time series. Modern approaches such as CNN and LSTM models are considered. We evaluated our prediction based on the MAE metric (mean absolute error). From the obtained results, we can conclude that it is appropriate using deep neural networks in predicting player behavior.

Keywords – CNN, LSTM, neural networks, player behavior prediction, time-series forecasting

I. INTRODUCTION

In the last decade, online gaming has become one of the most popular sources of entertainment. The gaming industry accounts for a significant portion of the market. The global online gambling market is anticipated to be worth more than \$92.9 billion by 2023. The market is currently valued at almost \$59 billion, and it is expected to double in size over the next several years. Unfortunately, gaming has several severe socioeconomic consequences. Furthermore, there is evidence that some types of gambling can lead to pathological gambling (also known as compulsive gambling or disordered gambling) and suicide[1]. With the introduction of online cyber gambling in the mid-1990s, the overall effect and prevalence of positive contributions and negative issues increased. Cyber gambling makes gambling significantly more widespread in society, significantly loosening restrictions on the time and place of gambling.

This combination, specifically the gambling industry, is already a multi-billion-dollar sector that is rapidly developing, owing in part to the widespread availability of cyber-gambling sites in recent decades. There is a matter of balancing between gambling contribution and social well-being.

The benefits and drawbacks of gambling are dependent on the participants' actions and judgments, particularly their bets. On the one hand, gambling's contribution to the gambling industry's income, employment, and government revenues is directly dependent on the players' bets and losses, as well as gambling-related problems such as financial pain, pathological gambling, and possible suicide.

Gambler betting models can aid the work of relevant government authorities, gambling sector operators, economists, sociologists, psychiatrists, and psychologists in the face of this dispute. Such models, in particular, assist public policymakers in reformulating critical public policies and proposing new legislation to regulate the gaming business to provide adequate government revenue while resolving gambling-related issues. Such models also enable gaming operators and their staff to reposition and develop their games in order to maximize earnings and, as a result, provide job security. At the same time, economists and sociologists can better assess the socioeconomic value of the gaming industry based on the inferences made by such models regarding betting amounts, while psychiatrists and psychologists can better determine the behavior and decisions of players and improve the treatment and rehabilitation of pathological players.

II. CURRENT APPROACHES TOWARDS MODELING PLAYER BEHAVIOUR

The problem to which this paper is devoted is the modeling of player behavior. Such a problem is closely related to the topic of time-series forecasting. There are various methods of time series forecasting, which will be analyzed in the following sections.

A. Time-series forecasting

Time-series forecasting (depicted on Figure 1) is the process of examining time series data and making forecasts and strategic decisions using statistics and modeling. It is not always an accurate prediction, and the likelihood of making one varies greatly especially when dealing with rapidly shifting variables in time-series data and events outside our control. Nonetheless, forecasting can show some events are more likely or less likely than others. The more information we have, the more accurate our prediction will be. Although forecasting and predicting are often used interchangeably, there is a distinction to be made. Forecasting may refer to data at a specific future point in time in some businesses, whereas predicting refers to future data in general. Time-series analysis and forecasting are frequently used together. Time-series analysis is creating models to comprehend the data and determine the underlying reasons. Analysis enables you to comprehend the "why" underlying the observed results. The next step in predicting is deciding what to do with that knowledge and extrapolations of what might happen in the future[2].

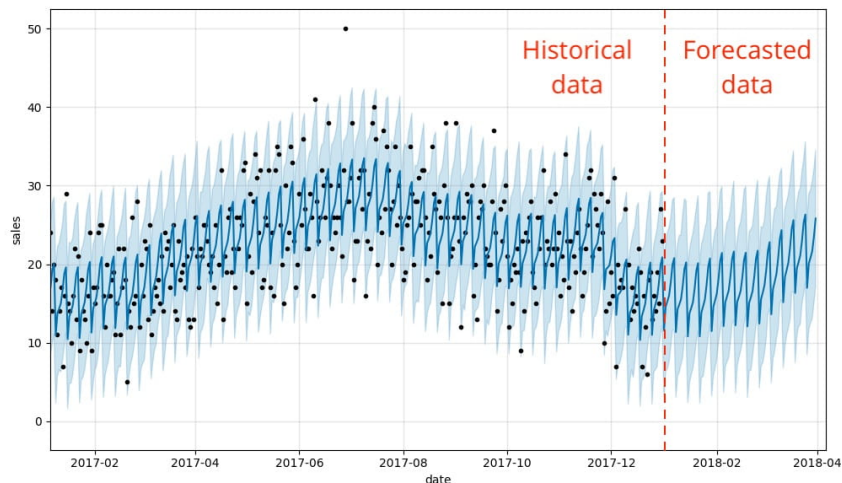


Fig. 1 Time-series forecasting visualization [3].

B. Theory-driven approaches

Experience shows that while theoretical approaches can make sense of and justify model choices, such models often lack the necessary functions because their architects lack insight. Games, being essentially open-ended constructions, tend to create a vast space of action. For this reason, data-driven approaches are promising because they do not rely on domain expertise[4]. In addition to relying less on expert guidance, data-driven approaches can also collect a massive amount of low-level user actions before any modular description of those actions.

C. Statistical approaches

In [5], the authors question the superiority of using deep neural networks in predicting time-series. The goal of the study was to demonstrate the capabilities of a set of different machine learning methods compared to classical statistical time-series prediction methods on a very large and diverse collection of univariate time-series prediction problems. The problem with this article is that it does not consider more complex time-series problems, where there are complex relationships between several variables. Moreover, the model must account for more than one characteristic, for the player bet prediction problem described above. For example, a player might have a strategy based on the symbols that appeared on a reel in a slot game, or he might adjust his bet based on his overall account balance.

D. Deep learning approaches

Traditional approaches can be overly simplistic and generally require labor-intensive pre- and post-processing steps to ensure satisfactory performance. Meanwhile, deep learning models have made great strides over the past few decades. Their application in natural language processing (NLP) and image classification has consistently outperformed traditional approaches and disrupted business as usual. As for time-series prediction, many research papers have successfully applied deep learning techniques[6]. They have proposed models that can overcome the challenges posed

by statistical approaches, and better handle the complexity of time-series forecasting and thus obtain significantly better results.

1) *Multi layer perceptron*: The multilayer perceptron (MLP) is an extension of the forward neural network. Mechanism of the algorithm is described in [7]. Reasons for using MLP in time-series forecasting problems are as follows:

- Resistance to noise. Neural networks are robust to noise in the input data and mapping function, and can even support learning and prediction in the presence of missing values.
- Nonlinearity. Neural networks make no strict assumptions about the mapping function and easily learn linear and nonlinear relationships.

The application of MLP for sequence prediction requires that the input sequence be divided into smaller overlapping subsequences, which are presented to the network to create the prediction. The time steps of the input sequence become the input features for the network. The subsequences overlap, simulating a window that shifts along the sequence to generate the desired output. This method may work well for some problems, but it has four critical limitations[8]:

- Stateless. MLPs are trained to approximate a fixed function. Any outputs that depend on the context of the input sequence must be generalized and frozen in the weights of the network.
- Ignorance of temporal structure. Temporal steps are modeled as input features, which means that the network has no explicit processing or understanding of temporal structure or order between observations.
- The inputs are fixed in size. The sliding window size is fixed and must be superimposed on all network inputs.
- The outputs are fixed in size. The output size is also set, and all outputs that do not match it must be forced.

2) *Convolutional neural networks*: Convolutional neural networks have their origins in image processing. It was first published in LeNet for handwritten digit recognition MNIST [9]. However, convolutional neural networks are not limited to image processing. CNNs convolutions are popularly known to work on spatial or 2D data. What's less popular is that there are also convolutions for 1D data. This allows CNN to be used in more general data type including texts and other time-series data. Instead of extracting spatial information, 1D convolutions can be used to extract information along the time dimension[10].

Here are some advantages of CNN:

- CNN has no suggestion that the story is complete, CNNs learn patterns within the time window. If there is missing data, CNNs should be useful.
- Feature learning. Automatic identification, extraction and distillation of meaningful features from raw data that relate directly to the modeled prediction problem.
- CNNs are computationally cheaper than such types of neural networks as RNNs, that will be described later. CNNs are trained in batches, while RNNs are trained sequentially. Thus, the RNN cannot use parallelization, since it must wait for previous calculations.

CNNs gain the advantages of multilayer perceptrons for time-series prediction, namely support for multivariate input, multivariate output, and learning arbitrary but complex functional relationships, but do not require the model to learn directly from lag observations. Instead, the model can learn a representation from a large input sequence that is most relevant to the forecasting problem.

3) *Recurrent neural networks*: A recurrent neural network (RNN) is a type of artificial neural network that uses sequential or time-series data [11]. These deep learning algorithms are typically used for ordinal or temporal tasks such as language translation, natural language processing (nlp), speech recognition, and image captioning; they are included in popular applications such as Siri, voice search, and Google Translate[12]. Recurrent neural networks, like feedforward and convolutional neural networks (CNNs), use training data for learning. They differ in their "memory" because they take information from previous inputs to influence the current input and output.

The main problem with CNNs and MLPs is that they do not remember what they have learned, i.e., after each iteration of the network learning, it starts all over again without remembering anything it learned during the previous iteration. This property puts them at a disadvantage when the data is sequential.

4) *LSTM*: LSTM (Long Short-Term Memory) is a Recurrent Neural Network (RNN) based architecture that is widely used in natural language processing and time-series forecasting [13]. After a prediction is made, it returns to the model to predict the next value in the sequence. With each prediction, some error is introduced into the model. To avoid exploding gradients, the values are "compressed" using (usually) sigmoid and tangent activation functions before the input and output gates.

III. EXPERIMENTS

One of the key factors that was taken into account to select the right features was the domain specific knowledge about slot game players. Following features were selected: MONEY - Absolute account balance of an individual player. WINS - win from game round. STAKES - A stake of the player for a game round. This feature will also be treated as the one we want to predict. TIME - The time in seconds that has passed between previous and current round. SYMBOLS - Represent a string with the symbols that player sees on a reel bands during game round.

Dataset consisted of 1 million game rounds played by 10 000 different players. Players were separated by their ID number.

A. Preprocessing the data

Preprocessing techniques can be applied.

- 1) ID can be removed, since it has no sufficient information for the model and will only hinder of the model training.
- 2) Only the individual sequence of players should be passed into the model. One problem with the players in the dataset is that they are all gathered together and have different numbers of games played. Therefore, without properly separating the players, we may pass the end of one player's sequence and the beginning of another player's sequence into the model, which is undesirable.
- 3) Symbols encoding. The representation of symbols in the game is a string. The problem is that the model accepts only numbers, so a tokenization technique should be applied to convert string to number. For this task, every combination of symbols will have its number.
- 4) Data normalization. MinMax scaling is applied to normalize the dataset.
- 5) Batching. This step is usually performed in order to speed up the learning process.
- 6) Split data to three parts: train data, validation data, test data.

B. Data windowing

In order to work with time-series data in supervised learning fashion we must transform them. By obtaining a sequence of numbers for the time-series data set, the data can be reconstructed so that it looks like a supervised learning problem. It can be done by using the previous time steps as input variables and using the next time step as the output variable.

The main features of the input windows are:

- 1) The width (number of time steps) of the input and label windows.
- 2) Which features are used as inputs, labels, or both.

Depending on the task and type of model, it is possible to generate a variety of data windows. Here are some examples:

- Multi step window for supervised learning.

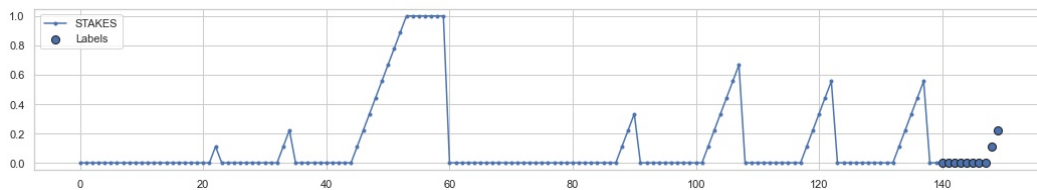


Fig. 2 140 player stakes given, 10 next stakes should be forecasted.

- Single step window for supervised learning.

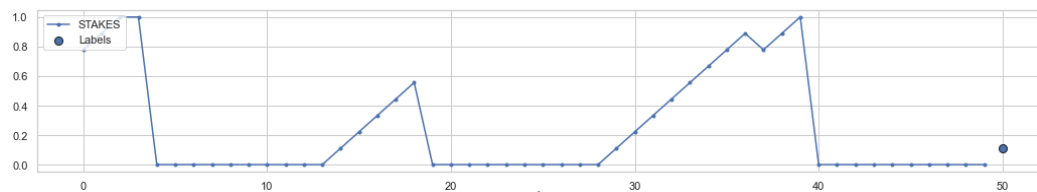


Fig. 3 140 player stakes given, 1 next stake should be forecasted.

C. Training the models

Based on the research, three types of models were tested: MLP, CNN, and LSTM. There are a couple of ways to train models:

- Try a different data generation approach.
- Try different window sizes.
- Choose models with a different structure.
- Tune hyperparameters.

There are some basic hyperparameters that can be configured: LEARNING_RATE = 3e-3; INPUT_WINDOW_SIZE = 50; PREDICTION_WINDOW_SIZE = 10; DROP_FEATURES = [] - Some features may be dropped; FEATURES = 5; LABELS = ['STAKES']; BATCH_SIZE = 64; TRAIN_VAL_TEST_FRAQ = (0.7, 0.2, 0.1)

Test data are used to evaluate the model, which means that the model is evaluated on data it did not see during training. The key evaluation metric is the mean absolute error (MAE), and is calculated using the following formula [14]:

$$MAE = \frac{\sum_{i=1}^n |y_i - x_i|}{n}$$

Mean absolute error is a model evaluation metric used with regression models. The mean absolute error of a model for a test set is the mean of the absolute values of the individual prediction errors over all instances in the test set.

Different types of models are used with different hyperparameters. Experiments were divided into two categories:

- Prediction based on long-term dependencies.
- Prediction based on short-term dependencies.

D. Overall results

In total, we did three experiments. There was no change in the hyperparameters described above in the first test. We removed some features that may not impact player stakes in the second experiment and modified hyperparameters. We also changed the way data were generated. Lastly, we tested the models for extreme predictions. There were two tests. First, we created a large window to find a long-term dependence on the players bets. If the model works well in this case, there is some continuing strategy in the players bets. In the second test, we created a narrower window to answer whether very short turns in the past affect player choices.

Table 1
 Overall results among all three experiments

Model	MAE	Input_Size	Output_Size	Experiment_N	Features_Count
MLP	0.42	50	10	1	5
CNN	0.43	50	10	1	5
LSTM	0.36	50	10	1	5
MLP	0.13	30	10	2	3
CNN	0.14	30	10	2	3
LSTM	0.11	30	10	2	3
MLP	0.19	140	10	3	3
CNN	0.24	140	10	3	3
LSTM	0.15	140	10	3	3
MLP	0.074	5	5	3	3
CNN	0.073	5	5	3	3
LSTM	0.068	5	5	3	3

As can be seen in Table 1, the model performs better if we consider a smaller amount of previous steps, i.e., slot game players decide how much to stake based on the small number of previous steps. In general, achieved results are satisfactory. The experiments show that models can predict the trend in player stakes far enough to reach the goals described in the introduction.

IV. CONCLUSION

As the result of this work, the overview of approaches toward time-series forecasting problems is described. Further, we focused on deep neural networks, especially LSTM and CNN, widely used in other fields where machine learning is applied. A dataset was created that can be used in similar problems related to predicting player behavior (for example, you can predict the time

between turns). The main outputs of this work are trained models that can predict player bets to the future. Three different types of models were created. We can conclude that the convolutional neural network has the worst performance based on the MAE metric. Multilayer perceptron has an average performance. A recurrent neural network with LSTM cells has the best performance over other networks since it has the best fit for the selected problem.

During the evaluation of the results, it is derived that slot gamblers make their bet decisions based on small amount previous rounds and have no long term strategy.

Nevertheless, there are some open questions or improvements left. Data that are passed to the model have the form of specific values instead of a relative number. For example, every win or stake of the player can be transformed into a percentage of the gain and the losses. This means we transfer from the regression problem into a classification problem. We can perceive a player's behavior in categories, e.g., the player will raise the bet or lower it, or continue with the same bet. So we can classify a set of some previous player bets into a specific category (raise a bet in the future, lowers a bet in the future). The approach described can provide a more general explanation of player behavior instead of a raw number that we see as an output of the regression.

REFERENCES

- [1] W. Feigelman, B. S. Gorman, and H. Lesieur, "Examining the relationship between at-risk gambling and suicidality in a national representative sample of young adults," *Suicide and Life-Threatening Behavior*, vol. 36, no. 4, pp. 396–408, 2006. [Online]. Available: <https://guilfordjournals.com/doi/abs/10.1521/suli.2006.36.4.396>
- [2] Tableau, "Time series forecasting: Definition, applications, and examples." [Online]. Available: <https://www.tableau.com/learn/articles/time-series-forecasting>
- [3] Databricks, "Fine-grained time series forecasting." [Online]. Available: <https://databricks.com/blog/2020/01/27/time-series-forecasting-prophet-spark.html>
- [4] D. Hooshyar, M. Yousefi, and H. Lim, "Data-driven approaches to game player modeling: A systematic literature review," *ACM Comput. Surv.*, vol. 50, no. 6, jan 2018. [Online]. Available: <https://doi.org/10.1145/3145814>
- [5] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, "Statistical and machine learning forecasting methods: Concerns and ways forward," *PLOS ONE*, vol. 13, no. 3, pp. 1–26, 03 2018. [Online]. Available: <https://doi.org/10.1371/journal.pone.0194889>
- [6] J. F. Torres, D. Hadjout, A. Sebaa, F. Martinez-Ivarez, and A. Troncoso, "Deep learning for time series forecasting: A survey," *Big Data*, vol. 9, no. 1, pp. 3–21, 2021, pMID: 33275484. [Online]. Available: <https://doi.org/10.1089/big.2020.0159>
- [7] M. Gardner and S. Dorling, "Artificial neural networks (the multilayer perceptron)a review of applications in the atmospheric sciences," *Atmospheric Environment*, vol. 32, no. 14, pp. 2627–2636, 1998. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1352231097004470>
- [8] J. Brownlee, *Deep Learning for Time Series Forecasting: Predict the Future with MLPs, CNNs and LSTMs in Python*. Machine Learning Mastery, 2018. [Online]. Available: <https://books.google.sk/books?id=o5qnDwAAQBAJ>
- [9] "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [10] B. Zhao, H. Lu, S. Chen, J. Liu, and D. Wu, "Convolutional neural networks for time series classification," *Journal of Systems Engineering and Electronics*, vol. 28, no. 1, pp. 162–169, 2017.
- [11] D. E. Rumelhart and J. L. McClelland, *Learning Internal Representations by Error Propagation*, 1987, pp. 318–362.
- [12] Y. Wu, M. Schuster, Z. Chen, Q. V. Le, M. Norouzi, W. Macherey, M. Krikun, Y. Cao, Q. Gao, K. Macherey, J. Klingner, A. Shah, M. Johnson, X. Liu, . Kaiser, S. Gouws, Y. Kato, T. Kudo, H. Kazawa, K. Stevens, G. Kurian, N. Patil, W. Wang, C. Young, J. Smith, J. Riesa, A. Rudnick, O. Vinyals, G. Corrado, M. Hughes, and J. Dean, "Google's neural machine translation system: Bridging the gap between human and machine translation," 2016. [Online]. Available: <https://arxiv.org/abs/1609.08144>
- [13] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [14] C. J. Willmott and K. Matsuura, "Advantages of the mean absolute error (mae) over the root mean square error (rmse) in assessing average model performance," *Climate Research*, vol. 30, pp. 79–82, 2005.

Analýza Nehodovosti na Cestách v USA Pomocou Vhodných Metód Dátovej Analytiky

Michal MICHŇÁK, Viera ANDERKOVÁ, František BABIČ

Katedra kybernetiky a umelej inteligencie, Fakulta elektrotechniky a informatiky,
Technická univerzita v Košiciach, Letná 9, 042 01 Košice, Slovenská republika

michal.michnak@student.tuke.sk, viera.anderkova@tuke.sk, frantisek.babic@tuke.sk

Abstrakt — Článok prezentuje analýzu dátovej vzorky o dopravných zaznamenaných na území USA v rokoch 2016 až 2020. Cieľom bolo identifikovať kľúčové faktory, ktoré spôsobujú takéto dopravné nehody a poskytnúť kompetentným orgánom zaujímavé znalosti potrebné pre rozhodnutia zamerané na zvýšenie bezpečnosti alebo zlepšenie kvality cestných komunikácií. Na tento účel sme využili vybrané metódy prediktívneho (Naivný Bayesovský klasifikátor, k-najbližších susedov, logistická regresia a náhodný les) a popisného dolovania (Atribútovo-orientovaná indukcia). Na základe výsledkov analýzy súčasného stavu si dovoľíme konštatovať, že AOI ešte nebola aplikovaná na takýto typ dát. Táto metóda funguje na princípe mazania atribútov či hierarchickom postupe naprieč atribútmi. Najlepšie výsledky pri klasifikácii dosiahli Naivný Bayes a logistická regresia. AOI nám umožnila postupne generalizovať pomerne veľkú vstupnú vzorku dát na charakteristické vzorky pre jednotlivé typy dopravných nehôd.

Kľúčové slová — Atribútovo-orientovaná indukcia, dátová analytika, dopravné nehody

I. ÚVOD

V súčasnosti je zníženie nehodovosti na cestách problémom, s ktorých sa snažia vysporiadať kompetentné orgány v krajinách naprieč celým svetom [1]. Približne 1.3 milióna ľudí ročne zomrie pri dopravných nehodách, čo predstavuje číslo nadmieru vysoké a vzbudzujúce pozornosť. Straty na životoch spôsobujú veľké ekonomické problémy jednotlivcom, ich rodinám a krajine. Tieto následky dopravných nehôd môžu vplývať na zvyšujúce sa náklady na liečbu či stratu hospodárskej produkcie jednotlivcom po smrti, či pri trvalom zranení. V dôsledku dopravných nehôd klesá hrubý domáci produkt väčšiny krajín o celé 3%. Problematike dopravných nehôd sa venuje široké spektrum ľudí so spoločným cieľom a to znížiť počet úmrtí pri dopravných nehodách na minimum, resp. eliminovať počet dopravných nehôd a zistiť príčiny dopravných nehôd. Zistenie príčin dopravných nehôd je nevyhnutné pre optimalizáciu cestnej dopravy, umožnenie bezpečnejších ciest a efektívnejšieho cestovania, za účelom spraviť samotné cestovanie bezpečnejšími. Do tejto kategórie zaraďujeme aj cyklistov či chodcov, ktorých bezpečnosť je rovnako dôležitá pri posudzovaní dopravných nehôd.

Dôležitým poznatkom pri dopravných nehodách je aj krajina. Až 93% dopravných nehôd, ktoré končia smrťou, sa objavili na cestách v krajinách nízkej a strednej príjmovej triedy, a to pri fakte, že v týchto krajinách sa nachádza len 60% vozidiel sveta. Najviac smrteľných zranení bolo zaznamenaných na Africkom kontinente. Ďalším zaujímavým faktom je používanie mobilných zariadení počas šoférovania. Pravdepodobnosť, že šofér používajúci mobilný telefón bude účastníkom nehody je štvornásobne vyššia ako šofér, ktorý sa plne venuje riadeniu motorového vozidla. Taktiež používanie prilby na motocykloch a bicykloch vedie k zvýšeniu bezpečnosti a zníženiu pravdepodobnosti fatálnych zranení až o 42% a zranení hlavy o 69%. Pátranie po ďalších príčinách dopravných nehôd je témou od začiatku 21. storočia, a je nevyhnutné, aby táto problematika bola naďalej skúmaná za cieľom k dosiahnutiu bezpečnejších ciest na celom svete.

A. Analýza Súčasného Stavú

Výskumné práce môžeme klasifikovať do dvoch samostatných skupín. Prvou je skupina prác, kde sú primárne analyzované vplyvy prírodných faktorov či životného prostredia ako je počasie, vlastnosti prostredia a cestnej siete a plynulosť premávky. Autori článkov [2]–[5] využívajú analýzu dát na nájdenie spojení a súvislostí pri predpovedí závažnosti nehôd. Eisenberg [6]

analyzoval príčiny autonehôd na pôde Spojených štátov v rokoch 1975-2000. Sústredil sa na zohľadnenie vplyvu počasia na závažnosť zranení spôsobených pri autonehodách a ich súvislosti ako množstvo a čas množstva zrážok.

Druhá skupina prác je zameraná na predpoveď frekvencie nehôd a klasifikáciu dopravných nehôd. Ide o predikciu približného počtu nehôd podľa informácií na vstupe o podmienkach cesty a iných udalosti ovplyvňujúcich jazdu na cestách. Ren a kol. [7] sa zamerali na identifikáciu súvislosti medzi dopravnými nehodami pomocou LSTM neurónových sietí. Výsledky použitých metód nedokázali odzrkadliť dôležité faktory, ktoré mohli mať vysoký vplyv na nehodu ako plynulosť premávky, ľudský faktor, typ cesty a jej vlastnosti či iné udalosti s vplyvom na situáciu na ceste, a preto vyprodukované mapy rizika boli značne odlišné od tej reálnej. Shokouhyar a kol. [8] využili dáta o dopravných nehodách na území Iránu. Za týmto účelom použili zhľukovanie, klasifikácia a extrakcia pravidiel. Najlepší dosiahnutý výsledok sa pohyboval na úrovni 76,32%. Krishnaveni a kol. [9] sa orientovali na porovnanie účinnosti klasifikačných modelov pri predikcii závažnosti nehôd. Pričom využili algoritmy ako Naivný Bayes, AdaBoost, PART klasifikátor, rozhodovací strom J48 a náhodné lesy. Výsledky práce ukázali, že najúčinnjší bol náhodný les, ktorý pri klasifikácii vozidiel či pozadia dokázal vykázat presnosť na úrovni 89,81% resp. 95,14%. AlMamlook a kol. [10] sa zamerali na vytvorenie modelu na správne selektovanie faktorov, ktoré majú závažný vplyv na výslednú vážnosť dopravnej nehody, a zostrojenie modelu pre klasifikovanie závažnosti zranení spôsobených pri dopravných nehodách. Autori využili viaceré algoritmy strojového učenia ako AdaBoost, logistická regresia, Naive Bayes a náhodný les. Výsledky tejto štúdie vykázali, že najlepšou voľbou pre predpovedanie závažnosti zranení pri dopravných nehodách bol náhodný les. Tento algoritmus vykázal úspešnosť 75,5%, logistická regresia dosiahla úspešnosť 74,5%, Naive Bayes vykázal vyše 73% a metóda AdaBoost skončila s presnosťou 74,5%. Umer a kol. [11] porovnávali stromové algoritmy (náhodný les, AdaBoost, Extra tree, Gradient Boost) s regresívnymi algoritmi (logistická regresia) pri predikcii závažnosti zranení spôsobených dopravnými nehodami. Štúdia potvrdila náhodný les ako algoritmus s najväčšou nameranou úspešnosťou, a to na úrovni 97,4%.

Počas analyzovania existujúcich výskumov sme zistili, že dopravné nehody boli riešené pomocou rôznych metód strojového učenia a mnohé z nich dosahovali vysoké úspešnosti. Pri hľadaní podobných prác sme narazili na fakt, že metóda atribútovo-orientovanej indukcie ešte nebola aplikovaná na dáta o dopravných nehodách.

II. NÁVRH VLASTNÉHO RIEŠENIA

Celý analytický proces dátovej vzorky sme realizovali pomocou metodológie CRISP-DM v programovacom jazyku R.

A. Pochopenie Cieľa

Primárnym cieľom nášho riešenia bolo prispieť k zlepšeniu dopravnej situácie naprieč všetkými svetadielmi a zvýšiť bezpečnosť na cestách. Na dosiahnutie tohto cieľa bolo potrebné zistiť príčiny dopravných nehôd a identifikovať, kedy sú dopravné nehody smrteľné. V rámci analýzy podobných prác sme zistili, že úspešnosť mnohých algoritmov a metód dosahovala v rámci vybraných metrík veľmi vysoké hodnoty (90+%). My sa pokúsime konkurovať podobným štúdiám a aplikovať metódy, ktoré doteraz neboli využité v kombinácii s dátami o dopravných nehodách. Ako cieľový atribút sme určili atribút Severity, ktorý popisuje závažnosť dopravnej nehody na dopravnú situáciu, spôsobené zranenia či zdržanie na ceste.

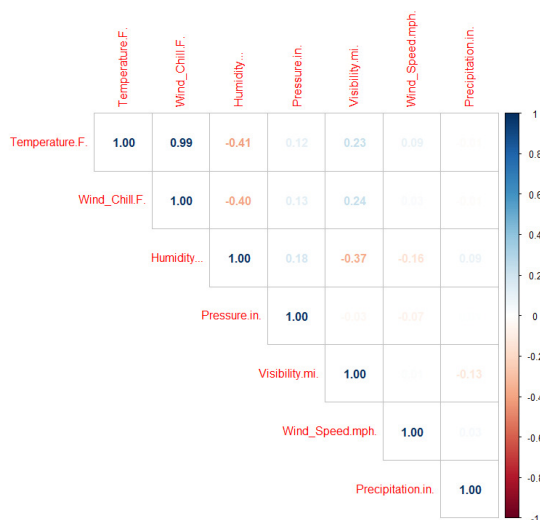
Ako primárne kritérium úspešnosti pri klasifikácii sme zvolili AUC. Pri realizácii atribútovo-orientovanej indukcie sme sa zamerali na identifikáciu všeobecných vzorov nehôd prostredníctvom postupnej generalizácie pôvodného počtu vyše 1,5 milióna záznamov.

B. Pochopenie a Príprava Dát

Pracovali sme s verejne dostupnou dátovou množinou US Accidents (updated) – A countrywide Traffic Accident Dataset (2016 – 2020) z webovej stránky Kaggle¹, ktorá slúži komunite vedcov v oblasti dát a odborníkov v oblasti strojového učenia. Vybraná dátová množina obsahuje vyše 1.5 milióna záznamov o dopravných nehodách z územia Spojených štátov Amerických, kde každý záznam bol popísaný 47 atribútmi, ktoré popisovali stav okolností s možným vplyvom na príčinu nehody.

Pri korelačnej analýze numerických atribútov (Obr. 1) sme identifikovali silnú závislosť medzi atribútom Temperature a atribútom Wind_Chill. Tieto atribúty vyjadrujú teplotu v stupňoch Fahrenheita. Zvyšné korelácie nedosiahli hodnoty ktoré by sme mohli považovať za dostatočné.

¹ <https://www.kaggle.com/sobhanmoosavi/us-accidents>



Obr. 1 Výsledok korelačnej analýzy.

Pomocou Pearsonovho Chi-kvadrát testu sme skúmali závislosť medzi nominálnymi atribútmi (Obr. 2). Skupina atribútov, vyjadrujúca či v momente nehody bol deň alebo noc (Sunrise_Sunset, Civil_Twilight, Nautical_Twilight a Astronomical_Twilight), vykázala v danom teste hodnoty p nižšie ako stanovaná hranica významnosti 0.05 (p -value < $2.2e-16$). To znamená, že dané atribúty medzi sebou vykazujú vysokú mieru asociácie, preto v ďalšej fáze budeme pracovať len s jedným atribútom. Následne sme tento test vykonali aj medzi atribútmi Amenity, Bump, Crossing, Give_Way, Junction, No_Exit, Railway, Roundabout, Station, Stop, Traffic_Calming, Traffic_Signal. Mnoho kombinácií nám rovnako vykázalo hodnoty p menšie ako 0.05. Na základe tejto analýzy sme zistili, že v kroku modelovania budeme môcť vynechať atribút Amenity (opisujúci prítomnosť nákupných centier, či iných budov s vysokou intenzitou pohybu ľudí), nakoľko takéto objekty sú takmer vždy kombinované inými atribútmi.

	Amenity	Bump	Crossing	Give_Way	Junction	No_Exit	Railway	Roundabout	Station	Stop	Traffic_Calming	Traffic_Signal
Amenity		9.008e-09	2.2e-16	1.62e-13	2.2e-16	2.2e-16	2.2e-16	0.8976	2.2e-16	2.2e-16	2.2e-16	2.2e-16
Bump			2.2e-16	0.9357	0.001724	5.36e-05	2.2e-16	1	6.506e-09	2.2e-16	2.2e-16	0.01229
Crossing				2.2e-16	2.2e-16	2.2e-16	2.2e-16	1	2.2e-16	2.2e-16	2.2e-16	2.2e-16
Give_Way					2.2e-16	0.0004174	0.21	1	0.09277	2.2e-16	0.6601	2.2e-16
Junction						0.2389	2.2e-16	2.2e-16	2.2e-16	2.2e-16	0.0007208	2.2e-16
No_Exit							2.034e-08	1	2.2e-16	2.2e-16	0.009246	2.2e-16
Railway								1	2.2e-16	2.2e-16	1.94e-07	2.2e-16
Roundabout									1	1	1	0.02057
Station										1	2.2e-16	2.2e-16
Stop												2.2e-16
Traffic_Calming												
Traffic_Signal												

Obr. 2 Výsledky Pearsonovho Chi-kvadrát test pre binárne atribúty.

Ďalším dôležitým krokom bolo vytvorenie nového atribútu s názvom Severity_Binary, ktorý obsahoval hodnotu „0“, ak v danom zázname o nehode bola uvedená vážnosť nehody od 1 po 3. Atribút obsahoval hodnotu „1“, ak záznam o nehode mal vážnosť nehody uvedenú číslom „4“. Odstránili sme atribút Number, ktorý obsahoval enormné množstvo prázdnych hodnôt a neposkytoval pre nás podstatnú informáciu; a všetky záznamy s prázdnyimi hodnotami, aby sme eliminovali ich vplyv na prvé testy. Týmto krokom sme odstránili vyše 550 tisíc riadkov a začali pracovať so vzorkou 943 579 záznamov. Vymazali sme záznamy obsahujúce prázdne polia (alebo tie ktoré obsahovali len „medzeru“), záznamy o mestách či poštových smerovacích číslach s nízkym zastúpením. Odstránené boli aj hodnoty o počasí, najpočetnejšie v rámci atribútu Wind_Direction, hovoriacom o smere vetra, ktoré boli písané slovom, čo malo za následok enormné množstvo rôznych kombinácií slov a skratiek popisujúci rovnaký smer. Z dátovej vzorky o veľkosti približne 950 tisíc záznamov, sme týmto krokom dokázali zredukovať počet na vyše 700 tisíc záznamov. Procesom čistenia sa odstránili aj záznamy, ktoré obsahovali hodnotu pre atribút „Severity_Binary“ rovnú „1“, čím sa zmenil pomer v rámci atribútu. Tento pomer bol preto rozdielny pre testy na dátach pred a po čistení. Za učiacu vzorku sme použili 9 001 riadkovú dátovú množinu a za testovaciu vzorku 999 riadkovú dátovú množinu z dôvodu rýchlosti spúšťania metód pri analyzovaní desiatok atribútov. Množiny disponovali rozdelením atribútu Severity_Binary pomerom 95:5.

C. Modelovanie a Vyhodnotenie

V prvej iterácii experimentov sme sa zamerali na vybrané atribúty bez hĺbkových úprav a zmien za účelom otestovania potenciálu vybraných klasifikačných metód. V rámci druhej iterácie experimentov sme vykonali rovnaké testy na očistených dátach.

Na základe vykonaných experimentov sme dospeli k záveru, že aj napriek malej vzorke a jednoduchým úpravám, vybrané metódy dokážu pomerne efektívne predpovedať závažnosť nehody. Rozsah pôvodnej dátovej vzorky bol veľmi veľký na spracovanie dát na normálnom zariadení, keď metódy pri spracovávaní vyše 20 atribútov vyžadujú desiatky GB pamäte RAM. Dosiahnutie AUC nad úroveň 0.90 však hodnotíme pozitívne a vidíme potenciál v daných dátach pre budúce experimenty. Netreba zabúdať však na fakt, že model náhodného lesa síce dosiahol po očistení dát výslednú hodnotu AUC nad 0.90, avšak nedokázal správne identifikovať ani jeden záznam triedy „1“. Najväčší potenciál z aplikovaných metód nám z daných testov vykazujú naivný bayesovský klasifikátor a logistická regresia (Tab. 1).

Tab. 1 Porovnanie výkonnosti klasifikačných metód pred a po očistení dát

Metóda	AUC – pred očistením	AUC – po očistení
Naivný Bayesovský klasifikátor	0.8175	0.9104
Klasifikátor k-najbližších susedov	0.7946	0.8230
Logistická regresia	0.8667	0.8893
Náhodný les	0.8764	0.9084

Metóda AOI nám umožnila nájsť spojitosti medzi vážnosťou nehôd, frekvenciou a podmienok s vplyvom na vážnosť nehody. Tento proces sme aplikovali na prúd, v ktorom sme zahrnuli aj kategorické atribúty obsahujúce hodnoty „TRUE“ a „FALSE“. Prvotným krokom bolo vytvorenie hierarchií konceptov pre atribúty, ktoré to umožňovali. Následne sme postupovali predpísanými krokmi (Tab. 2):

- Odstrániť atribút obsahujúci veľké množstvo jedinečných hodnôt ak pre neho neexistuje žiadna hierarchia konceptov (Hierarchické koncepty môžu byť vyjadrené prostredníctvom iných atribútov. V danom prípade sa hierarchicky nižšie umiestnený atribút odstráni).
- Generalizácia atribútu na základe hierarchie konceptov, keď sa aktuálne hodnoty nahradia hodnotami na hierarchicky vyššej úrovni všeobecnosti podľa danej hierarchie vytvorenej pre konkrétny atribút.

Tab. 2 Postupnosť krokov a vývoja atribútovo orientovanej indukcie

Krok	Popis	Počet jedinečných záznamov	Počet generalizovaných záznamov
0	Vstupný stav	942 988	0
1	Nahradenie atribútov o počasí	594 843	154
2	Nahradenie atribútu „Hour“ za „Day_Time“	519 888	203
3	Nahradenie atribútu „Year“ za „Years“	519 572	203
4	Odstránenie „Timezone“	519 324	203
5	Odstránenie „City“	396 683	266
6	Odstránenie „Side“	350 919	506
7	Odstránenie „County“	172 177	958
8	Nahradenie atribútu „month“ za „Months“	130 477	1 671
9	Odstránenie „Years“	111 252	1 675
10	Nahradenie „Months“ za „Months_Group“	89 882	1 687
11	Odstránenie „State“	20 312	4 424
12	Odstránenie „Months_Group“	14 636	5 645
13	Odstránenie „Weekday“	4 151	35 092
14	Odstránenie „Day_Time“	1 077	119 810

V rámci tohto procesu sme objavili viacero zaujímavých poznatkov, nielen vo finálnom výpise, ale aj v niektorých krokoch. V kroku 6. sme objavili najpočetnejší záznam pre závažnú nehodu s výrazným nárastom pred zvyškom dátovej vzorky; v 10. kroku sme našli najpočetnejšie zastúpenie záznamov reprezentujúce udalosti v noci pri vysokej vlhkosti v mesiacoch apríl až september, prítomnosť len dvoch záznamov reprezentujúcich závažné nehody v zimných mesiacoch; v kroku 11 sme objavili, že aj napriek veľmi dobrej viditeľnosti, štandardnom tlaku či suchom stave vozovky a bezvetria, dochádza k vysokému množstvu nehôd a najpočetnejšou

skupinou sú záznamy o nehodách z pondelkového rána v mesiacoch apríl až september, za vysokej vlhkosti.

State	Severity_Binary	Months_Group	weekday	Teplo	Vlhkost	Tlak	viditeľnosť	Zrazky	veternosť	Day_Time	V1
FL	0	10-3	sobota	Teplo	veľmi vlhko	Standard	veľmi dobrá	Sucho	Bezvetrie	Noc	1687
FL	0	10-3	štvrtok	Teplo	veľmi vlhko	Standard	veľmi dobrá	Sucho	Bezvetrie	Noc	1641
CA	0	10-3	sobota	Teplo	Vlhko	Standard	veľmi dobrá	Sucho	Bezvetrie	Noc	1563
CA	0	10-3	utorok	Teplo	Sucho	Standard	veľmi dobrá	Sucho	Bezvetrie	Podvečer	1543
CA	0	10-3	piatok	Teplo	Sucho	Standard	veľmi dobrá	Sucho	Bezvetrie	Pobede	1537
FL	0	10-3	pondelok	Teplo	veľmi vlhko	Standard	veľmi dobrá	Sucho	Bezvetrie	Noc	1530
CA	0	10-3	štvrtok	Teplo	Vlhko	Standard	veľmi dobrá	Sucho	Bezvetrie	Podvečer	1521
CA	0	10-3	utorok	Teplo	Sucho	Standard	veľmi dobrá	Sucho	Bezvetrie	Pobede	1503
CA	0	10-3	piatok	Teplo	Sucho	Standard	veľmi dobrá	Sucho	Bezvetrie	Podvečer	1444
CA	0	10-3	streda	Teplo	Vlhko	Standard	veľmi dobrá	Sucho	Bezvetrie	Podvečer	1438

Obr. 3 Výpis najpočetnejších záznamov pre proces AOI.

III. ZÁVER

Naším cieľom bolo analyzovať príslušnú dátovú vzorku, porovnať výkonnosť vybraných klasifikačných metód strojového učenia a aplikovať atribútovo-orientovanú Najlepšie hodnoty AUC a zároveň správnu identifikáciu záznamov pre triedu „1“ dosiahli metódy Naivný Bayesovský klasifikátor a logistická regresia. Pomocou atribútovo-orientovanej indukcie sa nám podarilo identifikovať zaujímavé vzory dopravných nehôd; napríklad sme zistili, že najväčší počet nehôd sa stáva v noci pri vysokej vlhkosti a dobrých poveternostných podmienkach. Dosiahnuté výsledky vytvárajú vhodný základ pre návrh a implementáciu systému na podporu rozhodovania v tejto doméne, ktorý by umožnil kompetentným autoritám zvýšiť bezpečnosť a optimalizovať plynulosť dopravy.

POĎAKOVANIE

Táto práca vznikla vďaka podpore Agentúry na podporu výskumu a vývoja na základe grantov č. APVV-20-0232 a APVV-17-0550 a Vedeckej grantovej agentúry MŠVVaŠ SR v rámci grantu č. VEGA 1/0685/2.

ZOZNAM REFERENCIÍ

- [1] O. mondiale de la santé *et al.*, *World Report on Road Traffic Injury Prevention*. World Health Organization, 2004.
- [2] J. Abellán, G. López, and J. de Oña, "Analysis of traffic accident severity using Decision Rules via Decision Trees," *Expert Syst. Appl.*, vol. 40, no. 15, pp. 6047–6054, 2013, doi: <https://doi.org/10.1016/j.eswa.2013.05.027>.
- [3] D. Jaroszowski and T. McNamara, "The influence of rainfall on road accidents in urban areas: A weather radar approach," *Travel Behav. Soc.*, vol. 1, no. 1, pp. 15–21, 2014, doi: <https://doi.org/10.1016/j.tbs.2013.10.005>.
- [4] J. D. Tamerius, X. Zhou, R. Mantilla, and T. Greenfield-Huitt, "Precipitation Effects on Motor Vehicle Crashes Vary by Space, Time, and Environmental Conditions," *Weather. Clim. Soc.*, vol. 8, no. 4, pp. 399–407, 2016, doi: 10.1175/WCAS-D-16-0009.1.
- [5] A. Theofilatos, "Incorporating real-time traffic and weather data to explore road accident likelihood and severity in urban arterials," *J. Safety Res.*, vol. 61, pp. 9–21, 2017, doi: <https://doi.org/10.1016/j.jsr.2017.02.003>.
- [6] D. Eisenberg, "The mixed effects of precipitation on traffic crashes," *Accid. Anal. Prev.*, vol. 36, no. 4, pp. 637–647, 2004, doi: [https://doi.org/10.1016/S0001-4575\(03\)00085-X](https://doi.org/10.1016/S0001-4575(03)00085-X).
- [7] H. Ren, Y. Song, J. Wang, Y. Hu, and J. Lei, "A Deep Learning Approach to the Citywide Traffic Accident Risk Prediction," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, 2018, pp. 3346–3351, doi: 10.1109/ITSC.2018.8569437.
- [8] S. Shokouhyar, E. Taati, and S. Zolfaghari, "The Effect of Drivers' Demographic Characteristics on Road Accidents in Different Seasons Using Data Mining," *PROMET - Traffic&Transportation*, vol. 29, p. 555, 2017, doi: 10.7307/ptt.v29i6.2342.
- [9] K. .s and M. Hemalatha, "A Perspective Analysis of Traffic Accident using Data Mining Techniques," *Int. J. Comput. Appl.*, vol. 23, 2011, doi: 10.5120/2896-3788.
- [10] R. E. AlMamlook, K. M. Kwayu, M. R. Alkasisbeh, and A. A. Frefer, "Comparison of Machine Learning Algorithms for Predicting Traffic Accident Severity," in *2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT)*, 2019, pp. 272–276, doi: 10.1109/JEEIT.2019.8717393.
- [11] M. Umer, S. Sadiq, A. Ishaq, S. Ullah, N. Saher, and H. A. Madni, "Comparison Analysis of Tree Based and Ensembled Regression Algorithms for Traffic Accident Severity Prediction," *CoRR*, vol. abs/2010.1, 2020, [Online]. Available: <https://arxiv.org/abs/2010.14921>.

Použitie algoritmu inšpirovaného svätojánskymi muškami pre numerickú optimalizáciu

¹Tomáš MIŠČÍK, ²Marián MACH

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

¹tomas.miscik.2@student.tuke.sk, ²marian.mach@tuke.sk

Abstrakt — Článok sa zameriava na problematiku riešenia úloh numerickej optimalizácie pomocou metaheuristických algoritmov, špeciálne tých ktoré sú inšpirované prírodnými fenoménmi. Pozornosť je zameraná na chovanie svätojánskych mušiek a ich vzájomné priťahovanie na základe komunikácie pomocou svetelných signálov. Článok sa zaoberá modelovaním tohto priťahovania v podobe algoritmu svätojánskych mušiek, pričom okrem základnej verzie algoritmu uvádza aj dve jeho modifikácie ako príklad modifikácie aktualizovania polohy niektorých mušiek ako aj príklad parametrickej modifikácie. V rámci experimentovania bolo realizovaných niekoľko experimentov s uvedenými variantmi tohto algoritmu v rámci optimalizácie zvolenej testovacej funkcie.

Príušové slová — Ackleyho funkcia, algoritmus svätojánskych mušiek, metaheuristika, numerická optimalizácia, rojový algoritmus

I. ÚVOD

Príroda bola inšpiráciou pre zavedenie mnohých metaheuristických algoritmov. Na základe skúseností z prírody mnohé prírodné fenomény (napr. prirodzený výber a prežitie, rôzne vzory chovania, spôsoby komunikácie) sa stali hlavnou ideou v pozadí viacerých metaheuristických algoritmov, ktoré označujeme ako prírodne inšpirované algoritmy.

Jedným z takýchto fenoménov je komunikácia svätojánskych mušiek pomocou blikajúceho osvetlenia. Existuje viacero druhov svätojánskych mušiek s vlastnými vzormi blikania, tvoreného krátkymi zábleskami. Svetlo je pri nich produkované procesom nazývaným bioluminiscencia. Pri komunikácii odpovedajúca muška bude reagovať napodobňovaním rovnakého vzoru alebo reagovaním iným špecifickým vzorom. Takáto komunikácia sa používa napríklad na prilákanie ďalších mušiek – keďže mušky sú jednopohlavné, jedna muška priťahuje ďalšie bez ohľadu na ich pohlavie.

Tento článok sa zameriava na algoritmus svätojánskych mušiek FA (Firefly Algorithm). Jeho štruktúra je nasledovná – v druhej kapitole sa venuje popisu metódy, pričom pozornosť je venovaná základnej verzii algoritmu a vybraným rozšíreniam algoritmu. V tretej kapitole je uvedená séria experimentov s danými verziami algoritmu na úlohe hľadania riešenia globálneho minima vybranej optimalizačnej funkcie.

II. ALGORITMUS SVÄTOJÁNSKYCH MUŠIEK

A. Základný algoritmus svätojánskych mušiek

Ak považujeme prostredie svätojánskych mušiek za priestor všetkých možných kandidátov na riešenie problému numerickej optimalizácie a svätojánske mušky za jednotlivých takýchto kandidátov, potom príťažlivosť medzi muškami navzájom a ich pohyb smerom k sebe môže byť inšpiráciou pre optimalizačný algoritmus. Jas mušiek bude definovaný priestorom optimalizačnej funkcie – každá muška reprezentuje bod v mnohorozmernom priestore a hodnota optimalizačnej funkcie v tomto bode bude určovať jas svetla danej mušky. Čím bude hodnota funkcie lepšia v zmysle optimalizácie (väčšia pri maximalizácii a menšia pri minimalizácii optimalizačnej funkcie), tým vyšší jas bude mať svetlo príslušnej mušky. Situácia pri stretnutí dvoch mušiek je na obr. 1.



Obr. 1 Princíp pohybu mušiek (zdroj: www.frfly.com).

Pohyb mušiek bude obmedzený [1] – poloha jednej sa nezmení. Pre akékoľvek dve blikajúce mušky sa tá menej jasnejšia bude pohybovať smerom k jasnejšej muške. Veľkosť posunu závisí od atraktívnosti. Atraktívnosť závisí od jasnosti priťahujúcej mušky a od vzdialenosti medzi muškami – so zväčšujúcou sa vzdialenosťou atraktívnosť klesá. Hodnota atraktívnosti je definovaná vzťahom

$$\beta = \beta_0 e^{-\gamma r^2} \quad (1)$$

kde r je vzdialenosť medzi muškami, β_0 je hodnota príťažlivosti medzi muškami pri vzdialenosti $r = 0$ a γ je koeficient absorpcie svetla v médiu prostredia. Vzdialenosť medzi akýmikoľvek dvomi muškami i a j je

$$r_{i,j} = \left(\sum_{k=1}^d (x_{i,k} - x_{j,k})^2 \right)^{0.5} \quad (2)$$

kde $x_{i,k}$ je k -tou zložkou priestorovej súradnice i -tej mušky a d je rozmernosť priestoru v ktorom sa hľadajú riešenia.

Aby algoritmus nebol založený čisto na deterministickom využívaní známej informácie (polohy a jasnosti jednotlivých mušiek) ale obsahoval aj zložku skúmania neznámych častí priestoru, výsledná poloha presúvanej mušky nebude presne na spojnici dvoch mušiek ale ešte bude modifikovaná náhodnou perturbáciou – posunie sa o náhodnú vzdialenosť náhodným smerom.

Posun menej jasnejšej svätójánskej mušky i smerom k jasnejšej muške j sa vykoná ako zmena súradníc i -tej mušky podľa vzťahu

$$x_i = x_i + \beta_0 e^{-\gamma r^2} (x_j - x_i) + \alpha(\text{rand} - 0.5) \quad (3)$$

kde *rand* reprezentuje náhodné číslo z uniformnej distribúcie nad intervalom [0,1]. Prvý člen pravej strany je pôvodná hodnota súradnice ktorá bude modifikovaná, druhý člen reprezentuje deterministické posunutie smerom k jasnejšej muške a tretí člen predstavuje náhodnú perturbáciu výslednej polohy.

Algoritmus svätójánskych mušiek je navrhnutý ako rojový algoritmus používajúci populáciu svätójánskych mušiek. V rámci tohto roja je muška priťahovaná nielen jednou inou muškou ale môže byť priťahovaná viacerými inými muškami (ak všetky sú jasnejšie ako daná muška). Muška potom sekvenčne vykonáva viacero pohybov v priestore, každý z nich zodpovedá približovanie sa k niektorej z jasnejších mušiek. A týmto spôsobom sú menené pozície všetkých mušiek, pričom jednotlivé mušky sú posúvané v sekvenčnom poradí. Zmena pozície všetkých mušiek vytvára jednu epochu (alebo generáciu) algoritmu. Zvyčajne algoritmus realizuje vopred zadaný počet epoch s tým, že si pamätá najlepšie riešenie bez ohľadu na to, v ktorej epoche bolo nájdené.

B. Modifikácie algoritmu

V zásade existujú dve hlavné triedy modifikácií algoritmu svätójánskych mušiek. Jedna z týchto tried je parametrická modifikácia. V základnej podobe algoritmu sú parametre definované užívateľom ako konštanty alebo sa určujú určitým vopred definovaným spôsobom. Ich zmenené hodnoty alebo spôsob určovania patrí do tejto triedy modifikácií. Druhá trieda je tvorená novými mechanizmami aktualizácie polohy jednotlivých mušiek [2].

V základnej podobe algoritmu najjasnejšia muška nemení svoju polohu, pretože neexistuje žiadna iná ktorá by bola jasnejšia a teda ju priťahovala. Naopak, všetky ostatné mušky sa v podstate pohybujú smerom k nej. Jediný spôsob zlepšenia je situácia, keď nejaká iná muška po zmene svojej polohy sa stane jasnejšou ako doteraz najjasnejšia muška. V prípade, že poloha najjasnejšej mušky je v lokálnom extrémne optimalizovanej funkcii, a medzi týmto extrémom a polohou ostatných mušiek sa nenachádza lepšie riešenie, tak po nejakom čase všetky mušky sa stretnú v okolí daného lokálneho extrému.

Pre podporenie lepšieho skúmania rôznych častí priestoru sa preto umožňuje aj pohyb najjasnejšej mušky na nové pozície [3]. Tento pohyb je realizovaný podľa vzťahu

$$x_i = x_i + \alpha(\text{rand} - 0.5) \quad (4)$$

kde sa uvažuje iba náhodná perturbácia polohy mušky, pričom má rovnaký tvar ako perturbačný člen vzťahu (3) pre zmenu polohy ostatných mušiek.

Príkladom parametrickej modifikácie je zmena určovania váhy náhodnej zložky pohybu α . Často sa vychádza z úvahy, že nastavenie jednej statickej hodnoty nie je najvhodnejšie riešenie, pretože rozdielne štádiá hľadania riešenia môžu vyžadovať rozdielne používanie randomizácie pohybu mušiek. Na začiatku hľadania by mohlo byť užitočné povoliť algoritmu väčšiu mieru skúmania priestoru zatiaľ čo na konci už iba skúmanie malého okolia identifikovaného v procese hľadania. Preto sa často používa dynamické nastavenie pri ktorom sa hodnota parametra mení od vyššej hodnoty k hodnote nižšej. Príkladom môže byť [4] kde zmena je realizovaná podľa vzťahu

$$\alpha = c \left(d + e^{0.005(\text{Iter} - \text{MaxIter})} \right)^{-1} \quad (5)$$

kde *Iter* označuje aktuálnu epochu a *MaxIter* zase celkový počet epoch. Hodnoty parametrov *c* a *d* určujú štartovaciu a konečnú hodnotu meneného parametra, ktorý bude začínať o niečo menšou hodnotou ako je *c/d* a končiť bude na hodnote *c/(d+1)*.

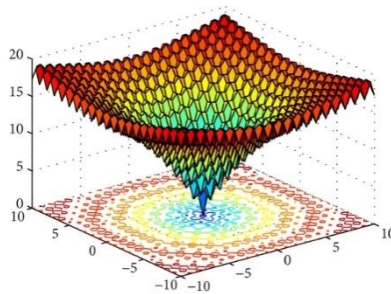
III. VYBRANÉ EXPERIMENTY

A. Testovací problém

Pre experimentovanie s algoritmom bola použitá pomerne známa Ackleyho funkcia [5], ktorá je často využívanou funkciou pre testovanie optimalizačných algoritmov. Táto funkcia je definovaná vzťahom

$$f(x) = -a \exp \left(-b \sqrt{\frac{1}{d} \sum_{i=1}^d x_i^2} \right) - \exp \left(\frac{1}{d} \sum_{i=1}^d \cos(cx_i) \right) + a + \exp(1) \quad (6)$$

kde pre jednotlivé konštanty boli použité typicky používané hodnoty $a=20$, $b=0.2$ a $c=2\pi$. Jej dvojrozmerná podoba ($d=2$) je zobrazená na obr. 2.



Obr. 2 Ackleyho funkcia dvoch premenných.

Globálny extrém tejto funkcie sa nachádza v počiatku súradnicového systému. Vo väčšej vzdialenosti od počiatku je funkcia charakterizovaná málo sa zvažujúcim tvarom, čo je kompenzované čoraz strmším poklesom pri približovaní sa ku globálnemu minimu. Periodická zložka, ktorá bola pridaná k základnému tvaru, vytvára veľké množstvo lokálnych extrémov, ktoré sťažujú prehľadávanie povrchu tejto funkcie. Voľbou $d > 2$ je možné získať viacrozmernú podobu funkcie, ktorá je tým obtiažnejšie optimalizovateľná, čím je použitý väčší počet dimenzií.

B. Počiatočné nastavenie parametrov

Pre algoritmus svätotajských mušiek je potrebné nastaviť štyri základné parametre – α (váha zohľadňujúca dôležitosť perturbačnej zložky), β_0 (príťažlivosť medzi muškami pri nulovej vzdialenosti medzi nimi), γ (koeficient absorpcie) a n (počet mušiek v roji). Postup nastavenia pozostával z dvoch fáz:

- voľba počiatočných hodnôt,
- iteračné zlepšovanie hodnôt.

V prvej fáze boli použité ako preddefinované hodnoty $\alpha = 1.0$, $\beta_0 = 1.0$, $\gamma = 0.01$ a $n = 10$. Tieto hodnoty boli použité pretože boli spomínané v rôznych literárnych zdrojoch.

V druhej fáze bol menený vždy iba jeden z parametrov zatiaľ čo hodnoty ostatných troch ostávali pevné, pričom po nastavení jedného parametra bol nastavovaný druhý atď. Pri tomto nastavovaní pre každú skúmanú hodnotu nastavovaného parametra bol algoritmus s danými hodnotami parametrov realizovaný 1000 krát, v každom behu po dobu 1000 epoch, pričom sa sledoval vývoj priemeru dosiahnutých riešení. Ako výsledná hodnota nastavovaného parametra bola použitá hodnota, ktorej zodpovedal najlepší priemerný výsledok. Priebeh nastavovania je v tab. 1.

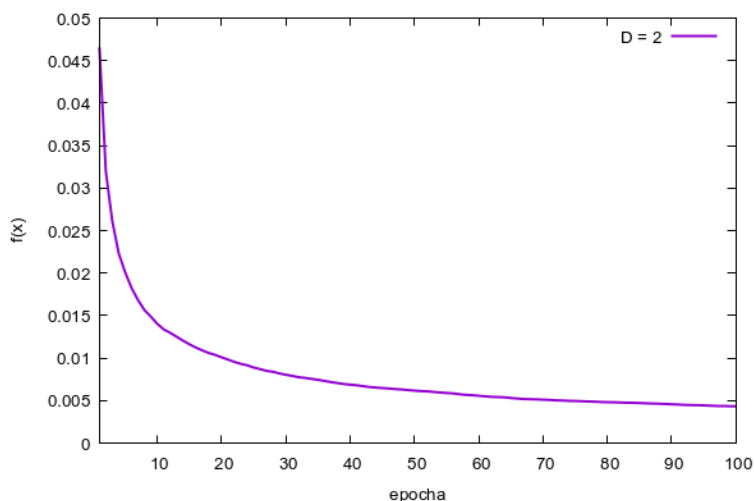
Tabuľka 1
 Postupné iteratívne nastavovanie parametrov riadené výsledkami pokusov s 1000-násobnou realizáciou algoritmu.

Poradie nastavenia	Nastavený parameter	Priemerné riešenie	Smerodajná odchýlka	Nastavená hodnota
		0.087497	0.056532	
1	α	0.004403	0.002283	0.06
2	γ	0.004274	0.002256	0.05
3	β_0	0.004184	0.002288	0.6
4	n	0.001425	0.000735	30

V tabuľke prvý riadok ukazuje dosiahnutý priemer riešení a súvisiacu smerodajnú odchýlku pre hodnoty parametrov stanovené v prvej fáze. Ostatné riadky ukazujú dosiahnutú kvalitu riešení pri vybranej hodnote parametra. Vidno, že najmenší vplyv na výsledok mal parameter γ . Pri určovaní počtu mušiek n pri ďalšom zväčšovaní hodnoty naďalej dochádzalo k zlepšovaniu priemerne dosahovaných riešení avšak s pomerne výrazným zvyšovaním časových nárokov. Preto bol tento počet obmedzený na zvolený počet mušiek.

C. Výsledky dosiahnuté základným algoritmom

Pre hodnoty parametrov, nastavené v predchádzajúcej kapitole, sme realizovali 1000 násobné opakovanie realizácie algoritmu na zvolenej testovacej funkcii so štatistickým spracovaním dosiahnutých výsledkov [6]. V každej epoche z dosiahnutých riešení bol určený priemer, smerodajná odchýlka a najlepšie riešenie spomedzi získaných riešení. Výsledky priemerných hodnôt pre prvých sto epoch sú zobrazené na obr. 3.



Obr. 3 Priebeh priemerného nájdeného riešenia

Ako vidno na obrázku, algoritmus pomerne rýchlo konverguje počas niekoľkých málo epoch. Následne nastáva postupné spomaľovanie poklesu hodnôt (hľadané globálne minimum má nulovú hodnotu), ktoré sa napokon ustáli na iba malej rýchlosti konvergencie k hľadanému globálnemu minimu. Otázkou je, dokedy krivka bude vykazovať (aj keď malý) pokles a kedy sa ustáli na nejakej hodnote. Hodnoty dosiahnuté v ďalších epochách sú v tab. 2.

Tabuľka 2
Vývoj konvergenzie základného algoritmu pri 1000-násobnom opakovaní.

Epocha	Najlepšie riešenie	Priemerné riešenie	Smerodajná odchýlka
500	0.000057	0.001955	0.001009
1500	0.000019	0.001101	0.000575
2500	0.000030	0.000838	0.000452
3500	0.000048	0.000726	0.000375
4500	0.000026	0.000624	0.000333
5500	0.000014	0.000582	0.000287

Je zrejmé, že približovanie sa ku globálnemu minimu poskytuje daný algoritmus iba v obmedzenej miere – najlepšie nájdené riešenie je presné iba na 4 desatinné miesta, priemerne sa algoritmus k svojmu cieľu približuje na 3 desatinné miesta.

D. Výsledky dosiahnuté modifikovaným algoritmom

Prvou modifikáciou bolo pridanie pohybu aj pre najjasnejšiu mušku podľa vzťahu (4). Pri tejto modifikácii všetky parametra ostali nastavené ako v predošlej časti. Dosiahnuté výsledky sú v tab. 3.

Tabuľka 3
Vývoj konvergenzie modifikovaného algoritmu (pohyb najjasnejšej mušky) pri 1000-násobnom opakovaní.

Epocha	Najlepšie riešenie	Priemerné riešenie	Smerodajná odchýlka
500	0.000113	0.000136	0.000136
1500	0.000038	0.000054	0.000045
2500	0.000027	0.000034	0.000045
3500	0.000028	0.000069	0.000063
4500	0.000013	0.000019	0.000043
5500	0.000026	0.000049	0.000042

Modifikovaná verzia algoritmu vykazuje stabilnejšie chovanie, o čom svedčia hodnoty smerodajnej odchýlky. Menšia variabilita nachádzaných riešení spolu s častejším nachádzaním lepších riešení sa zase prejavuje v rádomom znížení hodnoty priemeru nájdených riešení. Toto je však kombinované s faktom, že najlepšie nájdené riešenia sa nijako významne nelíšia od tých, ktoré boli nájdené základnou podobou algoritmu.

Druhou modifikáciou bola časovo sa meniaci váha náhodnej perturbácie pohybu mušiek podľa (5), pričom boli zvolené hodnoty $c=0.14$ a $d=2$ vďaka ktorým parameter α štartuje na hodnote 0.0698 a končí na hodnote 0.0467 (úmyselne bol zvolený takýto interval pretože obsahuje staticky nastavenú hodnotu pre daný parameter použitú v predchádzajúcich experimentoch). Dosiahnuté výsledky sú v tab. 4.

Tabuľka 4
Vývoj konvergenzie modifikovaného algoritmu (nelineárna zmena parametra) pri 1000-násobnom opakovaní.

Epocha	Najlepšie riešenie	Priemerné riešenie	Smerodajná odchýlka
500	0.000032	0.001935	0.001009
1500	0.000077	0.001219	0.000623
2500	0.000020	0.000933	0.000497
3500	0.000046	0.000824	0.000435
4500	0.000009	0.000734	0.000376
5500	0.000041	0.000664	0.000337

Pri porovnaní so základným algoritmom (tab. 2) je zrejmé, že pre danú testovaciu funkciu uvedená modifikácia nie je vhodná – a to ako z hľadiska priemeru nachádzaných riešení tak aj z hľadiska variability týchto riešení. Problém môže byť v nevhodne zvolených hraniciach pre

výber hodnoty parametra alebo v nelineárnom spôsobe zmeny. Preto sme zopakovali experiment so zmenou váhy perturbačnej zložky, keď sa váha menila v tých istých hraniciach ako predtým (od hodnoty 0.0698 až po hodnotu 0.0467) avšak lineárnym spôsobom. Dosiiahnuté výsledky sú v tab. 5.

Tabuľka 5
Vývoj konvergenzie modifikovaného algoritmu (lineárna zmena parametra) pri 1000-násobnom opakovaní.

Epocha	Najlepšie riešenie	Priemerné riešenie	Smerodajná odchýlka
500	0.000070	0.001842	0.000964
1500	0.000038	0.001107	0.000582
2500	0.000016	0.000806	0.000439
3500	0.000023	0.000685	0.000357
4500	0.000026	0.000611	0.000321
5500	0.000029	0.000556	0.000294

Ako vidno pri porovnaní s výsledkami základného algoritmu, výkonnosť algoritmu sa pridaním uvedenej modifikácie o niečo zvýšila. Ale táto modifikácia je očividne menej účinná než bolo pridanie pohybu najjasnejšej mušky. Pri skombinovaní oboch modifikácií boli získané výsledky, ktoré sú v tab. 6.

Tabuľka 6
Vývoj konvergenzie modifikovaného algoritmu (pohyb najjasnejšej mušky + lineárna zmena parametra) pri 1000-násobnom opakovaní.

Epocha	Najlepšie riešenie	Priemerné riešenie	Smerodajná odchýlka
500	0.000042	0.000123	0.000204
1500	0.000022	0.000034	0.000122
2500	0.000026	0.000056	0.000069
3500	0.000011	0.000020	0.000054
4500	0.000008	0.000017	0.000055
5500	0.000009	0.000030	0.000033

IV. ZÁVER

Vykonané experimenty potvrdili, že algoritmus založený na chovaní svätajánskych mušiek môže byť použitý ako optimalizačný nástroj v prípade riešenia úloh numerickej optimalizácie. Avšak je potrebné rátať s aproximátnym riešením – aj napriek veľkému počtu opakovaní experimentov sa algoritmu podarilo priblížiť ku skutočnému riešeniu so zhodou maximálne štyri desatinné miesta. Použitie modifikácií algoritmu malo pozitívny vplyv na jeho chovanie avšak hlavne v oblasti stability a variability dosahovaných riešení. Samotná kvalita najlepšieho nájdeného riešenia sa zlepšila iba o jeden rád, keď nájdené riešenie súhlasí so skutočným globálnym riešením v piatich rádoch. Na druhej strane otestovali sme iba dve modifikácie algoritmu, iné jeho modifikácie môžu mať potenciál väčšieho zlepšenia.

LITERATÚRA

- [1] X.S. Yang, *Nature-inspired metaheuristic algorithms*, Luniver Press, UK, 2010.
- [2] W.A. Khen et al., "A Review and Comparative Study of Firefly Algorithm and its Modified Versions", *Optimization Algorithms – Methods and Applications*, O.Baskan (ed), IntechOpen, 2016.
- [3] N. F. Johari et al., Firefly Algorithm for Optimization Problem, *Applied Mechanics and Materials*, roč. 421, 2013, 512-517.
- [4] Y. S. Yang a S. Su, Self-adaptive step firefly algorithm, *Journal of Applied Mathematics*, 2013.
- [5] T. Back: *Evolutionary Algorithms in Theory and Practice*, Oxford University Press, 1996, 142-143.
- [6] T.Miščík, *Použitie metód inšpirovaných svätajánskymi muškami pre riešenie úloh numerickej optimalizácie*, Bakalárska práca, Fakulta elektrotechniky a informatiky Technická univerzita v Košiciach, 2021.

Podpora Diagnostiky Diabetes Mellitus Pomocou Vhodných Metód Dátovej Analytiky

Natália ONUFRÁKOVÁ, Viera ANDERKOVÁ, František BABIČ

Katedra kybernetiky a umelej inteligencie, Fakulta elektrotechniky a informatiky,
Technická univerzita v Košiciach, Letná 9, 042 01 Košice, Slovenská republika

natalia.onufrakova@student.tuke.sk, viera.anderkova@tuke.sk,
frantisek.babic@tuke.sk

Abstrakt — Článok prezentuje analytický proces na podporu diagnostiky ochorenia Diabetes Mellitus. Na tento účel sme použili historické dáta o diabetes z Birla Institute of Technology, Mesra, vysokej školy v Jharkhand v Indii. Postupovali sme pomocou metodológie CRISP-DM a jednotlivé experimenty vytvárali pomocou programovacieho jazyka R. Naším cieľom bolo dosiahnuť efektívnejšie modely ako v existujúcich štúdiách, ktoré sme rozobrali v analýze súčasného stavu. Na predspracované dáta sme aplikovali šesť metód binárnej klasifikácie: Naivný Bayes, rozhodovací strom, SVM, náhodný les, logistická regresia a k-najbližších susedov. Na riešenie nevyváženej distribúcie hodnôt cieľového atribútu sme využili nadzorkovanie s rôznymi pomermi pri jednotlivých atribútoch. Keďže väčšina porovnávaných prác využívala metriku úspešnosť, využili sme ju aj my, ale primárne sme sa zamerali na metriky AUC a MCC.

KLúčové slová — diabetes mellitus, CRISP-DM, klasifikácia, kroková regresia

I. ÚVOD

Dátová analytika je v súčasnosti veľmi známa oblasť na získanie zaujímavých faktov z dátovej množiny a vo veľkej miere sa využíva vo finančníctve a marketingu [1]. Jednou z oblastí, v ktorej je dátová analytika čoraz viac populárnejšia a dôležitejšia, je práve zdravotníctvo. Zdravotníctvo produkuje obrovské množstvo dát, ktoré sú zložité a objemné na spracovanie. V rámci dátovej analytiky je možné využiť metodológiu na transformáciu týchto dát na užitočné informácie pri rozhodovaní lekárov a existuje obrovský potenciál pre aplikácie v zdravotnej starostlivosti. Môže zlepšiť rozhodovanie objavovaním vzorov vo veľkom množstve komplexných údajov, a tým pomôcť pri hodnotení účinnosti liečby, riadiť zdravotníctvo, či predikovať choroby.

V súvislosti s ochorením Diabetes mellitus každým rokom celosvetovo pribúda stále viac pacientov. Nakoľko je toto ochorenie skryté, až kým sa neprejaví vážnejšie zdravotné problémy s ním súvisiace, vzniká tak potreba toto ochorenie istým spôsobom zistiť skôr, než sa prejaví. Predikciou tohto ochorenia by ľudia mohli predísť zhoršeniu zdravotného stavu, dostať sa tak k zmene životného štýlu pomocou správnej životosprávy a vyhnúť sa tak komplikáciám či život ohrozujúcemu stavu.

Aby mohlo ľudské telo správne fungovať, potrebuje energiu, ktorú získava z prijatej potravy. Energiu si zabezpečujeme z cukrov, tukov a bielkovín. Cukry sa nachádzajú takmer vo všetkom, či už ide o ovocie, zeleninu, pečivo. V článku¹ sme sa dozvedeli, že na to, aby sa zložené cukry (polysacharidy) dokázali premeniť na energiu, musia sa v tráviacej sústave rozložiť na menšie časti a jednou z nich je práve glukóza, ktorá je základným zdrojom energie v tele človeka. V medicíne sa pre glukózu v krvi používa výraz glykémia. Glukózu v krvi je možné merať pomocou venóznej (žilovej) alebo kapilárnej krvi (z končekov prstov). Energia sa uchováva v pečeni a svaloch vo forme glykogénu ako energetická rezerva. Tá sa využíva počas noci alebo po cvičení.

K príznakom, ktoré pacient pociťuje pri zvýšenej hladine cukru v krvi patria: sucho v ústach a to najmä v nočných hodinách, slabosť, pokles hmotnosti, hnisavé infekcie kože, opakované

¹ <https://diabetik.sk/vsetko-o-diabetes-mellitus/zakladne-informacie-o-diabetes-mellitus-cukrovke/>

zápaly močových ciest, zhoršenie zraku či strata citlivosti prstov dolných končatín. Najzávažnejším prejavom diabetu (cukrovky) je upadnutie do kómy, plytké zrýchlené dýchanie spôsobené silným okyslením organizmu, stratou tekutín a vysokými hodnotami cukru v krvi.

A. Analýza Súčasného Stavú

Predikciou Diabetu mellitus sa zaoberalo viacero výskumníkov, avšak každý z nich sa na túto problematiku pozeral iným spôsobom.

Tigga a Garg [2] použili rôzne klasifikačné metódy strojového učenia na predikciu diabetu 2. typu. Výskumu sa zúčastnilo 952 respondentov vo veku 18 a viac rokov, z toho 580 mužov a 372 žien, ktorí odpovedali na 18 otázok týkajúcich sa zdravia, životného štýlu a rodinnej situácie. Odpovede boli zozbierané prostredníctvom online a offline dotazníka, z ktorých vznikol jednoduchá dátová množina². Autori sa zamerali na logistickú regresiu, k-najbližších susedov (K-NN), metódu podporných vektorov (SVM), Naivný bayes (NB), rozhodovací strom a náhodný les. Tieto algoritmy aplikovali aj na dátovú množinu Pima Indian Diabetes³ a výsledky porovnali. Výsledky ukázali, že dátová množina vytvorená pre tento výskum mala pri všetkých metódach vyššiu presnosť ako Pima Indian Diabetes dátová množina. Experimentálny výsledok ukázal, že úspešnosť náhodného lesa pri vytvorenej dátovej množine je 94,10%, čo je najvyššia hodnota oproti všetkým ostatným.

Han Wu a kol. [3] chceli v roku 2018 svojím výskumom vylepšiť presnosť predikcie diabetu a vytvoriť model, ktorý by sa dokázal prispôbiť na viac ako jedna dátová množina. V prvej úrovni bol použitý algoritmus K-means na odstránenie nesprávne zoskupených údajov a optimalizovaná množina údajov bola použitá pre klasifikáciu pomocou logistickej regresie. Na vyhodnotenie kvality modelu tiež použili ukazovatele ako úspešnosť, senzitivita, Matthewov korelačný koeficient (MCC), F1 skóre, ROC krivku a PRC krivku. Svojím výskumom dosiahli úspešnosť predikcie 95,461 %.

Cieľom analýzy [4] bolo nájsť optimálny klasifikátor, ktorý mal poskytnúť najbližší výsledok v porovnaní s klinickými výsledkami a pomohol by pri včasnej detekcii. V tejto štúdií boli využité metódy rozhodovací strom, NB, SVM, náhodný les a K-NN. Vylepšená úspešnosť Naivného Bayesa dosiahla hodnotu 82,30%, takže poskytla najlepšie prispôbenie údajom s ohľadom na diabetických a nediabetických pacientov.

Autori štúdie [5] použili na predikciu rozhodovací strom, náhodný les a neurónovú sieť. V tejto štúdií použili na meranie klasifikovanej účinnosti senzitivitu, špecificitu, úspešnosť a Matthewov korelačný koeficient. Použitý bol dataset Luzhou z údajov fyzického vyšetrenia nemocnice v Luzhou v Číne a Pima Indians, ktorý predstavuje údaje o diabetikoch Pima Indians. Najlepší výsledok pre množinu údajov Luzhou je 80,84% pri náhodnom lese a najlepší výsledok pre Pima Indians je 78,28% pri neurónovej sieti.

II. ANALYTICKÝ PROCES

Celý analytický proces bol realizovaný v súlade s metodológiou CRISP-DM, ktorá pozostáva zo šiestich fáz. Jednotlivé úlohy vo fázach metodológie CRISP-DM sme riešili pomocou voľne dostupného programovacieho prostredia RStudio v jazyku R.

A. Pochopenie Cieľa

V rámci biznis cieľa sme sa zamerali na zlepšenie presnosti diagnostiky ochorenia cukrovky o 5% oproti spomenutým výskumom. Diabetes je často skrytým ochorením a pacient o ňom nevie, pokiaľ sa neobjavia závažnejšie zdravotné problémy, ktoré s cukrovkou úzko súvisia. Vzniká tak potreba zaoberať sa predikciou tohto ochorenia a tak predísť zlému zdravotnému stavu pacienta.

Z pohľadu dolovania v dátach sme riešili binárnu klasifikáciu - ako cieľový atribút bol určený atribút Diabetic, ktorý obsahoval odpovede áno a nie, teda či opýtaný trpí alebo netrpí diabetom. Pre klasifikáciu boli použité algoritmy Naivný Bayes, rozhodovací strom, metóda podporných vektorov, náhodný les, logistická regresia a k-najbližších susedov.

B. Pochopenie a Príprava Dát

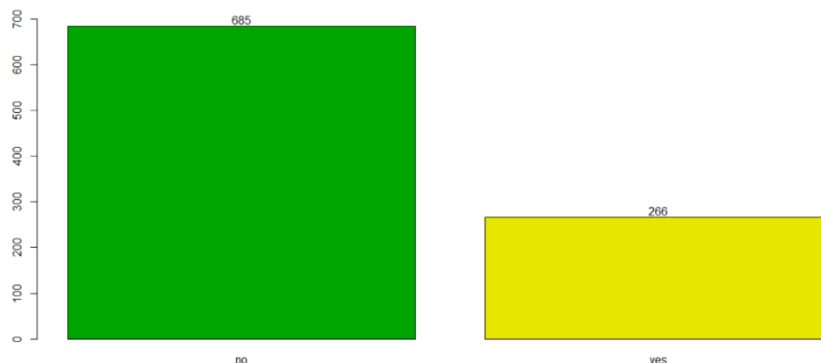
Dáta² využívané v tejto práci pochádzajú z Birla Institute of Technology, Mesra, vysokej školy v Jharkhand v Indii. Dáta boli zozbierané formou online a offline dotazníka, pričom sa zúčastnilo 952 respondentov vo veku nad 18 rokov. Dataset obsahoval 17 nezávislých atribútov a 1 závislý - Diabetic a nevyskytujú sa v nich žiadne osobné údaje účastníkov. Atribút Diabetic poskytuje

² <https://www.kaggle.com/datasets/tigganeha4/diabetes-dataset-2019>

³ <https://www.kaggle.com/uciml/pima-indians-diabetes-database>

informáciu o tom, či opýtaný má alebo nemá diabetes, 685 opýtaných cukrovkou netrpí, naopak 266 áno (viď Obr. 1).

Keďže sa v dátach vyskytovali aj chýbajúce hodnoty (pod označením NA), bolo potrebné ich nájsť a určiť ich počet.



Obr. 1 Výskyt cukrovky u pacienta.

Pomocou korelogramu sme overili závislosť numerických atribútov a zistili sme, že medzi atribútmi sa nenachádza vysoká závislosť. Následne sme pomocou Pearsonovho Chi-kvadrát testu overili závislosť medzi nominálnymi atribútmi.

Pri vytváraní modelov v ďalšej fáze budeme používať všetky atribúty okrem Stress, ktorý bol odstránený z dôvodu veľkého počtu závislostí s inými atribútmi a tiež z dôvodu, že stres je jedným z rizikových faktorov diabetu. Pri atribúte RegularMedicine bola hodnota "o" zmenená na hodnotu "no". Atribút BPLlevel, ktorý hovorí o výške krvného tlaku, obsahoval hodnoty s gramatickými chybami - tie boli opravené na správne hodnoty.

Numerický atribút BMI obsahoval štyri chýbajúce hodnoty, ktoré boli nahradené strednou hodnotou. Atribút Pregancies bol premenovaný na Pregnancies a chýbajúce hodnoty v ňom boli nahradené najčastejšie sa vyskytujúcou hodnotou, ktorá bola vypočítaná pomocou vytvorenej funkcie calc_mode(). Najčastejšie sa vyskytovala hodnota 0 tehotenstiev. V atribúte Pdiabetes sa vyskytovala jedna chýbajúca hodnota, ktorú sme tiež nahradili najčastejšie sa vyskytujúcou hodnotou. Cieľový atribút Diabetic obsahoval jednu chýbajúcu hodnotu, ktorá bola tiež nahradená najčastejšie sa vyskytujúcou hodnotou "no" a jedna hodnota s gramatickou chybou bola takisto zaradená k hodnote "no". Pre jednoduchšie pochopenie výsledných matíc vo fáze modelovania bola pri cieľovom atribúte hodnota "no" nahradená nulou a hodnota "yes" bola nahradená jednotkou. Nakoľko numerické atribúty dosahovali slabé a stredné korelácie, čo znamená, že si neboli príliš podobné, žiaden z týchto atribútov nebol odstránený. V prípade nominálnych atribútov sme na základe výsledkov Pearsonovho Chi-kvadrát testu zistili, že sa medzi jednotlivými kombináciami atribútov nachádzajú závislosti. Z tohto dôvodu bol odstránený iba jeden atribút Stress, ktorý mal najväčší počet závislostí s inými atribútmi.

Po úprave dát boli na dátach použité dva typy regresíí. Najprv bola použitá kroková regresia dopredu a po nej kroková regresia dozadu. Tieto regresie boli použité s cieľom nájsť kľúčové atribúty z pohľadu cieľovej diagnostiky. Ako vyhodnocovanie kritérium bolo použité AIC (Akaike Information Criterion). Výsledkom v oboch prípadoch bola nasledovná množina deviatich atribútov: RegularMedicine, Age, Pdiabetes, Family_Diabetes, PhysicallyActive, BPLlevel, Smoking, Pregnancies, Gender.

C. Modelovanie a Vyhodnotenie

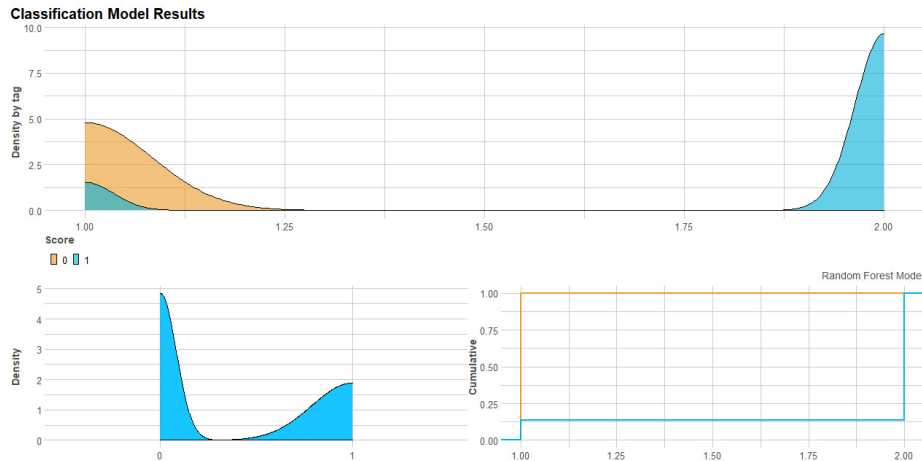
Pri vytváraní experimentov sme pracovali so závislou premennou Diabetic a 9 prediktormi, ktoré sme získali pomocou krokovej regresie. Pri prvom experimente sme vytvorili základné modely, bez úpravy dát. Dáta sme rozdelili na tréningovú a testovaciu množinu v pomere 80:20, neskôr v pomere 90:10. Sledovali sme viacero metrík, a to konkrétne úspešnosť, senzitivitu, špecifickosť, presnosť, F1 skóre, Matthewov korelačný koeficient, Kappa štatistiky a AUC plochu pod krivkou. Základnými experimentami sme dosiahli určité hodnoty sledovaných vyhodnocovacích metrík, tie sme však ďalej vylepšovali. Ako najlepšie sa zatiaľ javili algoritmy metóda podporných vektorov, náhodného lesa a logistickej regresie.

Následne sme rovnakú množinu algoritmov aplikovali na predpripravené data pri ďalších dvoch experimentoch a použili sme nadzorkovanie dát pomocou funkcie RandOverClassif() z dôvodu nevyváženej cieľovej množiny. Získali sme tak rovnomerné rozdelenie atribútu Diabetic pre 0 aj 1 pri rozdelení 80:20, rovnako aj pri rozdelení dátovej množiny v pomere 90:10. Pri každom algoritme bolo použité iné nadzorkovanie, ktorého úspešnosť sme vždy experimentálne overili.

Pre druhý experiment na rozdelenej dátovej množine 80:20 sme dosiahli najlepšie výsledky pri algoritmoch náhodného lesa, logistickej regresie a k-najbližších susedov. Algoritmus k-

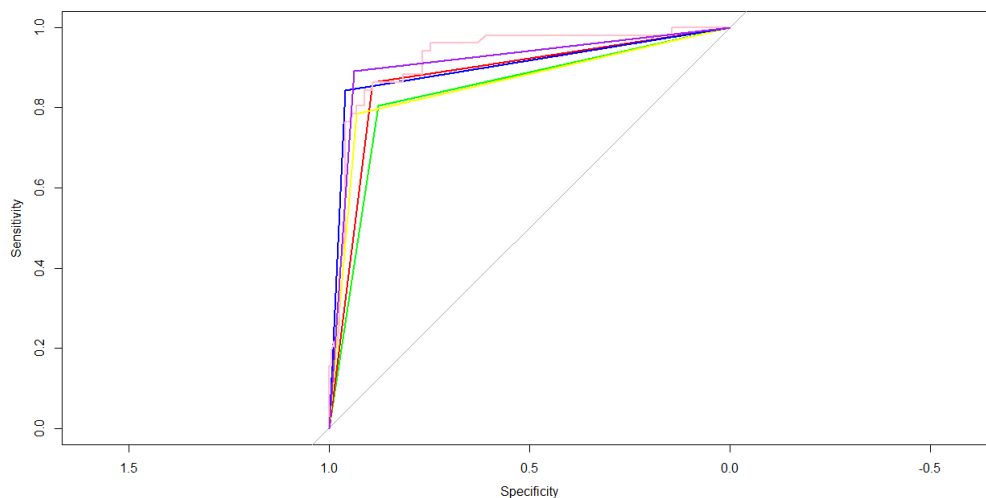
najbližších susedov dosiahol hodnoty MCC 81,44% a AUC krivky 91,38%. Druhým najlepším algoritmom pre predikciu diabetu bol model náhodného lesa s hodnotou MCC 81,26% a hodnotou AUC krivky 90,10%. Model logistickej regresie mal hodnotu MCC 75,49% a hodnota AUC krivky bola pri ňom rovná 92,57%. Tieto tri algoritmy sú teda na základe výsledkov z rozdelenia 80:20 najlepšie pre predikciu diabetu.

Výsledky klasifikačných modelov sme zobrazili v grafe hustoty (viď Obr. 2). Tento graf nám predstavuje vizuálne informácie o distribúcii a schopnosti nášho modelu rozlíšiť triedy. Vidíme teda, ako model rozdelil obe naše kategórie, čiže 0 a 1.



Obr. 2 Vyhodnotenie modelu vygenerovaného pomocou algoritmu náhodný les.

Krivka ROC na obrázku Obr. 3 ukazuje, ako model pracuje s našou testovacou množinou. Jednotlivé farby prislúchajú modelom: Naivný Bayes – zelená, rozhodovací strom – červená, metóda podporných vektorov – žltá, náhodný les – modrá, logistická regresia – ružová, k-najbližších susedov – fialová.



Obr. 3 ROC krivky modelov

Tretí experiment bol vykonaný na dátach rozdelených v pomere 90:10. Ako najlepšie algoritmy pre predikciu diabetu sa pri tomto rozdelení javili modely rozhodovacieho stromu, náhodného lesa a logistickej regresie. Najlepším z týchto modelov z pohľadu metriky MCC bol náhodný les s percentuálnou hodnotou 90,12% a hodnotou AUC krivky 96,30%. Druhým najlepším algoritmom z tohto hľadiska bol model logistickej regresie s hodnotou MCC 82,30% a hodnotou AUC krivky 95,50%. Ako tretí model sme uviedli rozhodovací strom, ktorý mal trochu menšiu hodnotu MCC, presne 80,06% a hodnotu AUC krivky pripadajúcu 93,45%.

Keďže iba prvá výskumná práca [2] použila vo svojom výskume rovnakú dátovú množinu ako my preto sme porovnali naše výsledky iba s výsledkami dosiahnutými pri tejto štúdií. V našej práci sme dosiahli úspešnosť 96,46% čo predstavuje zlepšenie o 2,36%.

III. ZÁVER

Diabetes mellitus je nevyspytateľné ochorenie, ktoré môže počas života nadobudnúť aj zdravo žijúci a stravujúci sa človek. Preto patrí výskum diabetes k populárnej oblasti skúmania a je veľmi

dôležitý v zdravotníctve aj kvôli častému výskytu v populácii a mnohým vyskytujúcim sa faktorom spojeným s touto diagnózou. Získané výsledky je možné využiť v zdravotníctve pri zisťovaní diabetu u pacientov počas preventívnych prehliadok. Bolo by však potrebné vytvoriť aplikáciu, ktorá by sa riadila našimi výslednými modelmi. Následne by bolo potrebné aplikáciu otestovať s lekármi, aby s ňou vedeli pracovať a naplno ju aj využiť. Ďalšou z možností, ako dosiahnuté výsledky využiť je vytvoriť kalkulačku diabetu, vo forme podobnej bežne dostupnej BMI kalkulačke na internete, ktorá by na základe vstupných údajov vedela predpovedať možnosť výskytu tohto ochorenia u človeka. Mala by viacero vstupných parametrov, po vložení ktorých by sa pomocou našich modelov vyrátalo nebezpečenstvo tohto ochorenia..

POĎAKOVANIE

Táto práca vznikla vďaka podpore Agentúry na podporu výskumu a vývoja na základe grantov č. APVV-20-0232 a APVV-17-0550 a Vedeckej grantovej agentúry MŠVVaŠ SR v rámci grantu č. VEGA 1/0685/2.

ZOZNAM REFERENCIÍ

- [1] H. C. Koh and G. Tan, "Data mining applications in healthcare.," *J. Healthc. Inf. Manag.*, vol. 19, no. 2, pp. 64–72, 2005.
- [2] N. P. Tigga and S. Garg, "Prediction of Type 2 Diabetes using Machine Learning Classification Methods," *Procedia Comput. Sci.*, vol. 167, pp. 706–716, 2020, doi: <https://doi.org/10.1016/j.procs.2020.03.336>.
- [3] H. Wu, S. Yang, Z. Huang, J. He, and X. Wang, "Type 2 diabetes mellitus prediction model based on data mining," *Informatics Med. Unlocked*, vol. 10, pp. 100–107, 2018.
- [4] N. Sneha and T. Gangil, "Analysis of diabetes mellitus for early prediction using optimal features selection," *J. Big data*, vol. 6, no. 1, pp. 1–19, 2019.
- [5] Q. Zou, K. Qu, Y. Luo, D. Yin, Y. Ju, and H. Tang, "Predicting diabetes mellitus with machine learning techniques," *Front. Genet.*, vol. 9, p. 515, 2018.

Návrh regulátorov a simulačné overenie výkonového multiport meniča

¹Adrián MARCINEK, ²Marek PÁSTOR

¹Katedra elektrotechniky a mechatroniky, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

²Katedra elektrotechniky a mechatroniky, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

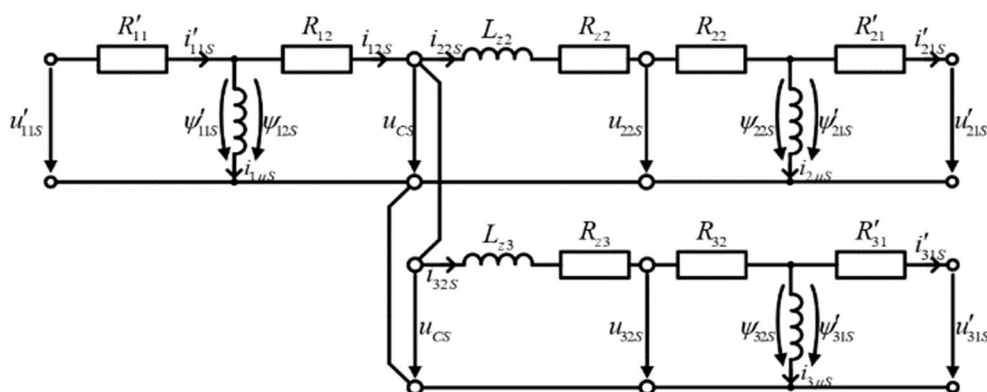
¹adrian.marcinek@tuke.sk, ²marek.pastor@tuke.sk

Abstrakt — Príspevok nadväzuje na publikáciu *Matematický model výkonového multiport meniča*, v ktorom je výkonový multiport menič popísaný matematickými rovnicami a uvedená je aj jeho náhradná schéma. V tomto príspevku sa budeme venovať návrhu regulátorov prúdov a tiež simulačnému overeniu riadenia pri konkrétnych podmienkach. Pri návrhu regulátorov a simulačnom overení je zachovaná topológia multiport meniča z predchádzajúceho príspevku a teda predstavíme výpočet regulátorov a simulačné overenie pre trojportový multiport menič, ktorého porty sú viazané pomocou dvojvinutových planárnych transformátorov.

Kľúčové slová — výkonový multiport menič, výpočet regulátorov multiport meniča, riadenie multiport meniča

I. ÚVOD

Príspevok priamo nadväzuje na publikáciu v zborníku *Electrical Engineering and Informatics* s názvom *Matematický model výkonového multiport meniča* [1], ktorý opisuje postup zostavenia matematického modelu takéhoto meniča na základe teórie riadenia elektrických pohonov. Uvažuje sa trojportový multiport menič, ktorý je napájaný harmonickým napätím. Jednotlivé porty v skutočnosti reprezentujú dvojvinutové planárne transformátory. Tieto transformátory sú prepojené elektricky pomocou ich sekundárnych vinutí. Vinutia sú prepojené paralelne a uzol prepojenia sekundárnych vinutí transformátorov nazývame ako „mediobvod“. Planárne transformátory sú použité kvôli vysokej objemovej hustote výkonu a modularite, ktorá predstavuje jednu zo základných požiadaviek. Keďže hovoríme o multiport meniči, je na jednotlivé porty meniča možné pripojiť ako zdroj tak aj spotrebič energie. Vo vyššie spomenutom príspevku je uvedená náhradná schéma multiport meniča ale aj samotný matematický model vo forme rovníc, ktoré sme využili ďalej pri návrhu regulátorov. Možné spôsoby riadenia, jeho použitie a samotný návrh je opísaný aj v [2] – [6]. V našom prípade sú rovnice matematického modelu najprv uvedené v α - β sústave, ktorú budeme nazývať „stojacou“. Veličiny v tejto sústave sú označené indexom „S“ a môžeme ich vidieť aj v náhradnej schéme multiport meniča na Obr.1. Ďalej sú rovnice transformované do x - y sústavy, ktorá rotuje uhlovou rýchlosťou odvodenou od frekvencie napätia resp. od spínacej frekvencie. Veličiny sú v tejto sústave označené indexom „K“ a pri rozložení do zložiek sú označené pomocou indexov y a x na činnú a jalovú zložku.



Obr. 1 Náhradná schéma multiport meniča [1]

Princíp riadenia výkonového multiport meniča bude nasledovný:

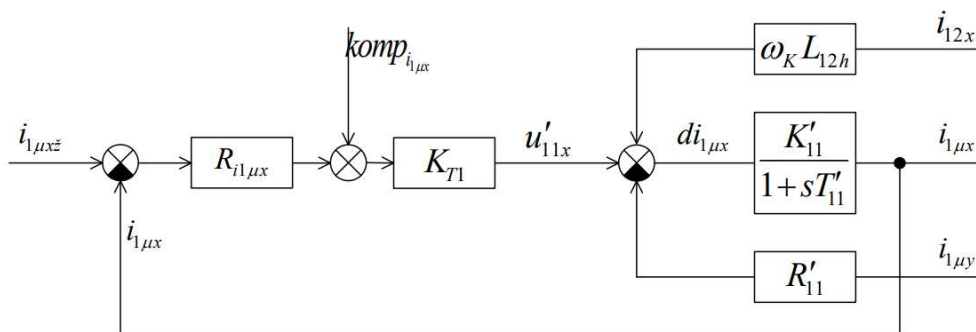
- regulátory umožnia nezávislé riadenie činnej a jalovej zložky prúdu – zložky prúdov y a x .
- výstupné napätie transformátora T_1 bude riadené magnetizačným prúdom $i_{1\mu}$. Rozdiel napätia medzi primárnou a sekundárnou stranou transformátora bude malý, keďže odpor vinutia a rozptylová indukčnosť vybraného planárneho transformátora sú taktiež malé. Indukované napätie na T_1 tak bude zodpovedať napätiu v medziobvode u_C . Prúd $i_{1\mu}$ bude riadený napätím u'_{11} .
- činné a jalové prúdy transformátorov T_2 a T_3 budú riadené regulátormi pre prúdy i_{22} resppektíve i_{32} . Riadiacou veličinou týchto prúdov budú napätia u'_{21} resppektíve u'_{31} . Regulátory i_{22} a i_{32} definujú činný prúd, ktorý je odoberaný, resp. dodávaný transformátormi T_2 a T_3 do medziobvodu. Rozdiel týchto prúdov dodá zdroj 1 cez transformátor T_1 .

II. NÁVRH REGULÁTOROV PRE MULTIPORT MENIČ

Ako bolo uvedené aj v predchádzajúcom príspevku, riadené budú prúdy $i_{1\mu}$, i_{22} a i_{32} . V nasledujúcej kapitole budú kvôli rozsiahlosti uvedené iba rovnice regulátorov pre jalové zložky prúdov $i_{1\mu}$, i_{22} a teda $i_{1\mu x}$, i_{22x} .

A. Návrh regulátora prúdu $i_{1\mu}$

Prúd $i_{1\mu}$ predstavuje magnetizačný prúd transformátora T_1 , jeho jalovú zložku budeme označovať $i_{1\mu x}$ a jeho činnú zložku $i_{1\mu y}$. Ďalej uvidíme návrh regulátora jalovej zložky magnetizačného prúdu od blokovej schémy až k výslednému prenosu s rovnicami konštant regulátora a činnú zložku magnetizačného prúdu iba ako rovnicu regulátora spolu s rovnicou kompenzačného signálu. Bloková schéma pre regulátor $i_{1\mu x}$ je na Obr. 2. K_{T1} predstavuje zosilnenie meniča a $R_{i1\mu x}$ predstavuje regulátor prúdu $i_{1\mu x}$. Pre jednoduchosť bol prenos prúdového snímača zvolený ako rovný jednej, preto nie je znázornený v blokovej schéme. Prúd $i_{1\mu x}$ je získaný výpočtom z meraných prúdov i_{11}' a i_{12} .



Obr. 2 Bloková schéma regulátora prúdu $i_{1\mu x}$

Ako zosilnenie a časovú konštantu primárneho vinutia zavedieme nasledujúce parametre

$$K'_{11} = \frac{1}{R'_{11}} \quad \text{a} \quad T'_{11} = \frac{L_{12}}{R'_{11}}. \quad (1)$$

Po úprave rovníc z predchádzajúceho príspevku dostaneme výslednú rovnicu jalovej zložky magnetizačného prúdu T_1 ako

$$i_{1\mu x} = \frac{K'_{11}}{1+sT'_{11}} (u'_{11x} - R'_{11}i_{12} + \omega_K L_{12h} i_{1\mu}), \quad (2)$$

kde prvý člen v zátvorke predstavuje zložku napätia na primárnom vinutí T_1 , druhý člen je zložka od záťažového prúdu a tretí člen predstavuje y – novú zložku magnetizačného prúdu. Pre výpočet regulátora ďalej predpokladáme, že všetky tieto veličiny sú merateľné. Pre zjednodušenie regulácie budeme druhý a tretí člen rovnice (2) kompenzovať. Takto nám ostane lineárna sústava, pre ktorú navrhujeme regulátor. Rovnica pre kompenzačný signál bude mať tvar

$$\text{komp}_{i_{1\mu x}} = \frac{1}{K_{T1}} (R'_{11}i_{12} - \omega_K L_{12} i_{1\mu}). \quad (3)$$

Pre návrh regulátora ďalej pokračujeme prenosom otvoreného obvodu

$$F_o = \frac{i_{1\mu x}}{i_{1\mu xž}} = R_{1\mu x} K_{T1} \frac{K'_{11}}{1+sT'_{11}}. \quad (4)$$

Regulátorom je možné nastaviť želanú dynamiku obvodu. Použitím PI regulátora zabezpečíme, aby výsledný prenos sústavy (aj s regulátorom) bol zotrvačnosťou prvého rádu s časovou konštantou T_{n1} . V takom prípade $T_{R1} = T'_{11}$. Prenos sústavy s regulátorom potom bude

$$\begin{aligned} F_o = \frac{i_{1\mu x}}{i_{1\mu xž}} &= K_{R1} \frac{1+sT_{R1}}{sT_{R1}} K_{T1} \frac{K'_{11}}{1+sT'_{11}} = K_{R1} \frac{1+sT'_{11}}{sT'_{11}} K_{T1} \frac{K'_{11}}{1+sT'_{11}} = \\ &= \frac{K_{R1} K_{T1} K'_{11}}{sT'_{11}} = \frac{1}{s \frac{T'_{11}}{K_{R1} K_{T1} K'_{11}}} = \frac{1}{sT_{n1}}. \end{aligned} \quad (5)$$

Kde T_{n1} predstavuje časovú konštantu výslednej sústavy. Ak je požiadavka, aby výsledná časová konštanta celého obvodu T_{n1} bola x -krát menšia ako pôvodná časová konštanta T'_{11} , potom využijeme rovnice

$$T_{n1} = \frac{T'_{11}}{K_{R1} K_{T1} K'_{11}} = \frac{T'_{11}}{x} \quad \text{a} \quad K_{R1} = \frac{x}{K_{T1} K'_{11}}. \quad (6)$$

V našom prípade bola požadovaná výsledná časová konštanta celého obvodu $T_{n1} = 0,005$ s, čo pre vyššie spomenutú premennú x z rovnice (6) predstavuje hodnotu $x = 11,342305$. Výsledná prenosová funkcia uzavretého systému pre riadenie prúdu $i_{1\mu x}$ je potom nasledovná:

$$F_W = \frac{i_{1\mu x}}{i_{1\mu xž}} = \frac{1}{1+s \frac{T'_{11}}{K_{R1} K_{T1} K'_{11}}} = \frac{1}{1+s \frac{T'_{11}}{x \frac{K_{T1} K'_{11}}{K_{R1} K'_{11}}}} = \frac{1}{1+s \frac{T'_{11}}{x}}. \quad (7)$$

Parametre simulačného modelu a regulátora prúdu $i_{1\mu x}$ sú uvedené v tabuľke Tab.1. Keďže uvažovaný výkonový multiport menič obsahuje tri rovnaké transformátory, parametre platia aj pre transformátory T_2 a T_3 . Parametre transformátorov použité v simulácii boli totožné s parametrami reálneho planárneho transformátora. Simulačný model multiport meniča bol zostavený v prostredí MATLAB/Simulink.

Tab. 1 Parametre simulačného modelu a regulátorov

Parameter	Hodnota
U_1	24 V
U_2	48 V
r_1	0,5

$R'_{11} = R_{12}$	0,0073 Ω
$L'_{11h} = L_{12h}$	13,5 μH
$L_{z2} = L_{z3}$	5 μH
$R_{z2} = R_{z3}$	0,05 Ω
$R_{c2} = R_{c3}$	0,064 Ω
$K_{T1} = K_{T2} = K_{T3}$	1
T'_{11}	0,0567 s
$T_{z2} = T_{z3}$	$7,7 \cdot 10^{-5}$ s
K'_{11}	137,15
$K_{c2} = K_{c3}$	15,484
$T_{n1} = T_{n2}$	0,005 s
T_{n3}	0,01 s
K_{R1}	0,0827
K_{R2}	$5 \cdot 10^{-4}$
K_{R3}	$1 \cdot 10^{-3}$

Pri výpočte regulátora $i_{1\mu y}$ vychádzame z mierne odlišnej rovnice

$$i_{1\mu y} = \frac{K'_{11}}{1+sT'_{11}} (u'_{11y} - R'_{11}i_{12y} - \omega_K L_{12h}i_{1\mu x}). \quad (8)$$

Postup návrhu regulátora vrátane parametrov regulácie z tabuľky Tab.1 je rovnaký ako pre jalovú zložku prúdu $i_{1\mu x}$. Návrh regulátora teda nebudeme uvádzať a uvedieme len kompenzačný signál pre tento regulátor.

$$\text{komp}_{i_{1\mu y}} = \frac{1}{K_{T1}} (R'_{11}i_{12x} + \omega_K L_{12h}i_{1\mu y}). \quad (9)$$

B. Návrh regulátora prúdu i_{22} a i_{32}

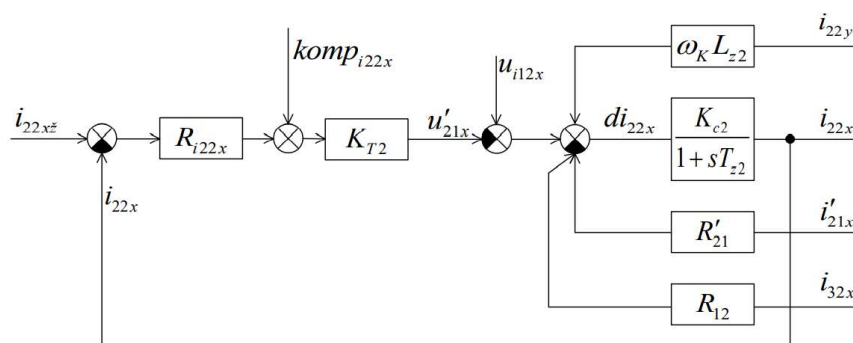
Princíp regulácie prúdov i_{22} a i_{32} je podobný ako pri $i_{1\mu}$, len sa pracuje s inými rovnicami. Pre názornosť uvedieme blokovú schému a samotný výpočet regulátora jalovej zložky prúdu i_{22x} . Pri výpočte regulátora vychádzame z rovnice:

$$i_{22x} = \frac{K_{c2}}{1+sT_{z2}} (u_{i12x} - u'_{21x} - R'_{21}i'_{21x} - R_{12}i_{32x} + \omega_K L_{z2}i_{22y}), \quad (10)$$

kde

$$K_{c2} = \frac{1}{R_{c2}} \quad \text{a} \quad T_{z2} = \frac{L_{z2}}{R_{c2}}. \quad (11)$$

Bloková schéma regulačného obvodu je na Obr. 3 Blok označený ako K_{T2} predstavuje prenos meniča a blok označený R_{i22x} predstavuje samotný regulátor prúdu i_{22x} . Prenos snímača prúdu predpokladáme rovný jednej, v schéme preto nevystupuje.



Obr. 3 Bloková schéma regulátora prúdu i_{22x}

V tomto prípade rovnako predpokladáme, že všetky veličiny v rovnici sú merateľné, preto nežiaduce väzby vieme kompenzovať signálom vyjadreným rovnicou (13). Zostane nám tak lineárna sústava, pre ktorú navrhne regulátor. Prenos otvoreného obvodu bude

$$F_o = \frac{i_{22x}}{i_{22z}} = R_{i22x} K_{T2} \frac{K_{C2}}{1+sT_{22}}. \quad (12)$$

Želanú dynamiku potom nastavíme pomocou PI regulátora. Parametre tohto regulátora spolu s parametrami obvodu sú taktiež uvedené v Tab. 1. Kompenzačný signál privedený za regulátor R_{i22x} je možné vyjadriť rovnicou

$$\text{komp}_{i_{22x}} = \frac{1}{K_{T2}} (u_{i12x} - R'_{21} i'_{21x} - R_{12} i_{32x} + \omega_K L_{z2} i_{22y}). \quad (13)$$

Polarita kompenzovaných veličín je daná voľbou polarity prúdu i_{21} v schéme. Princíp regulácie prúdu i_{22y} a tiež i_{32x} a i_{32y} je totožný, zmena nastane len v indexoch veličín v rovniciach, preto tieto výpočty ďalej uvádzať nebudeme.

III. SIMULAČNÉ OVERENIE MODELU S NAVRHNUTÝMI REGULÁTORMI

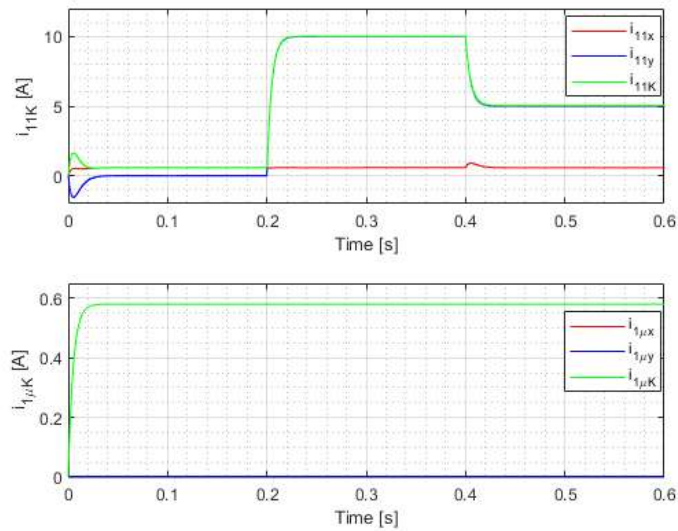
Riadenie multiport meniča bolo simulačne overené v programe MATLAB/Simulink. Pre simuláciu boli zvolené tri rovnaké planárne transformátory, ktorých parametre boli totožné s parametrami reálneho planárneho transformátora. Riadenie multiport meniča bolo simulačne overené v nasledujúcich situáciách:

- Počiatočné nastavenie magnetizačného prúdu – fázor magnetizačného prúdu $i_{1\mu k}$ bol reguláciou udržiavaný v osi x . Takto bolo zabezpečené konštantné napätie u_{i1} (a následne u_c). Magnetizačný prúd transformátora T_1 je dodaný z meniča do primárneho vinutia. Magnetizačné prúdy transformátorov T_2 a T_3 je potom možné nastaviť dvoma spôsobmi:
 1. nastavením jalovej zložky prúdu i_{22} . V tomto prípade je prúd dodaný z T_1 cez medziobvod. Tento režim je možné využiť, ak zdroje napájajúce T_2 a T_3 nemajú napätie a je nutné ho cez medziobvod nabiť,
 2. ak nežadáme žiadost' jalovej zložky prúdu i_{22} , napätie u_c prostredníctvom regulácie vyvolá magnetizačné prúdy $i_{2\mu k}$ pre T_2 a $i_{3\mu k}$ pre T_3 .
- Tok energie medzi portami – činné zložky prúdov transformátorov T_2 a T_3 sa nastavujú regulátormi R_{i22} a R_{i32} . Ak sa hodnoty prúdov i_{22x} a i_{32x} líšia, rozdiel prúdu dodá T_1 .

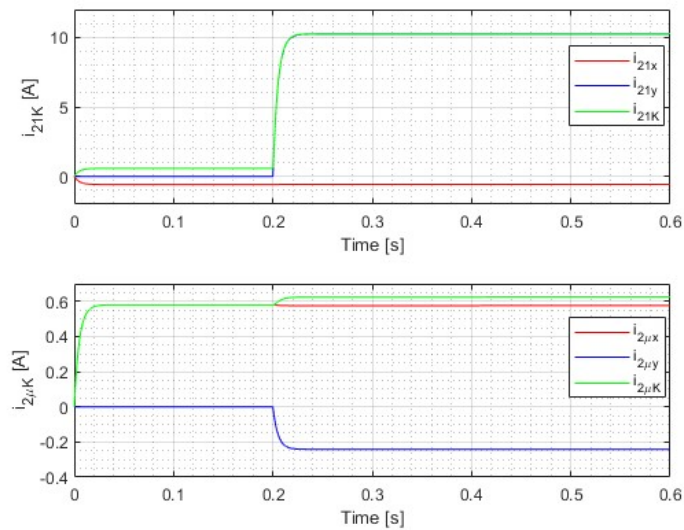
Obrázky Obr. 4, Obr. 5, Obr. 6 a Obr. 7 zobrazujú simulačné výsledky s nasledujúcim postupom:

- v čase $t = 0$ s bola zložka magnetizačného prúdu $i_{1\mu x}$ nastavená na hodnotu 0,58 A. Hodnota magnetizačného prúdu bola určená výpočtom na základe parametrov zvoleného planárneho transformátora. Magnetizačný prúd bol dodaný prostredníctvom meniča na primárnej strane transformátora T_1 ako jalová zložka prúdu i_{11} ,
- magnetizačné prúdy transformátorov T_2 a T_3 boli nastavené regulátormi prúdov i_{22} a i_{32} cez napätie medziobvodu u_c ,
- v čase $t = 0,2$ s, činná zložka prúdu i_{22} bola nastavená na hodnotu 10 A. To znamená, že zariadenia pripojené k T_2 odoberali daný prúd. Prúd bol dodávaný cez transformátor T_1 ,
- v čase $t = 0,4$ s, činná zložka prúdu i_{32} bola nastavená na hodnotu -5 A. To znamená, že zdroj pripojený k T_3 začal dodávať prúd. Prúd cez T_1 tak klesol na 5 A a transformátor T_1 dodával iba rozdiel prúdov medzi T_2 a T_3

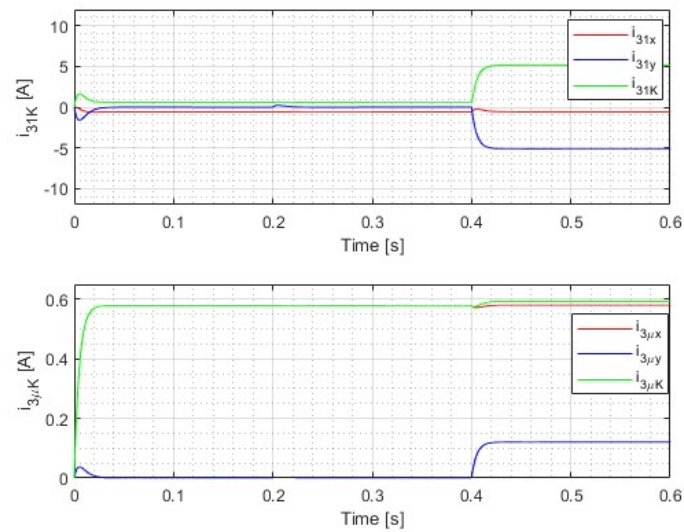
Na simulačných priebehoch vidíme, že regulácia činných a jalových zložiek prúdov je nezávislá.



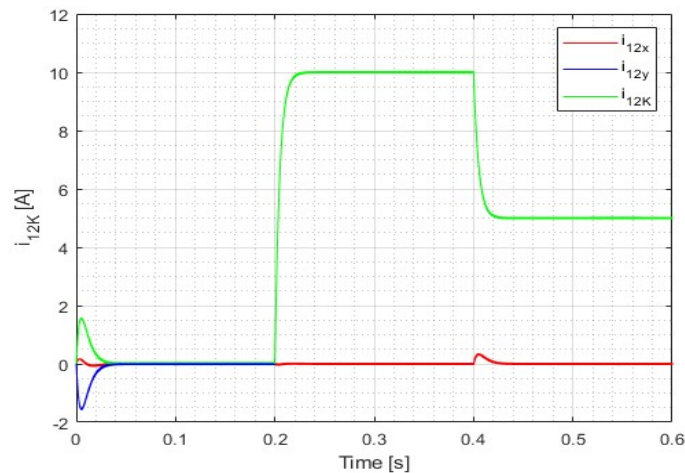
Obr. 4 Priebeh primárneho a magnetizačného prúdu transformátora T1



Obr. 5 Priebeh primárneho a magnetizačného prúdu transformátora T2



Obr. 6 Priebeh primárneho a magnetizačného prúdu transformátora T3

Obr. 7 Priebeh sekundárneho prúdu T_1 - prúd i_{12K} v medziobvode

Takto navrhnuté a simulačne overené riadenie multiport meniča je ďalej potrebné diskretizovať a upraviť tak, aby bolo vhodné pre implementáciu napr. do digitálneho signálového procesora. Do úvahy treba brať aj spínaciu frekvenciu, pri ktorej bol multiport menič simulovaný, tá predstavuje hodnotu 50 kHz. Takto vysoká spínacia frekvencia môže ovplyvniť meranie prúdu a teda aj samotné riadenie, preto je nutné, aby digitálny signálovým procesor disponoval dostatočne rýchlym AD prevodníkom pre meranie prúdu.

IV. ZÁVER

Príspevok popisuje postup návrhu a výpočet regulátorov jednotlivých veličín výkonového multiport meniča, ktorého matematický model spolu s náhradnou schémou boli vypracované v rámci predchádzajúcich príspevkov autora. Pri zostavovaní regulačnej schémy a výpočte regulátorov prúdu $i_{1\mu}$ sme predpokladali, že prúdy i_{11} a i_{12} sú merateľné a prúd $i_{1\mu x}$ sme získali výpočtom z meraných veličín. V prípade návrhu regulátorov prúdov i_{22} a i_{32} sme tiež predpokladali, že všetky veličiny vystupujúce v rovnici sú merateľné. Na základe týchto predpokladov mohli byť pre nežiadúce väzby zavedené kompenzačné signály. Tým sa zjednodušila regulačná schéma. S takto navrhnutou regulačnou schémou bol zostavený model multiport meniča v prostredí MATLAB/Simulink, kde bol multiport menič overený vyššie uvedeným postupom.

Pre reálne zhotovenie multiport meniča a implementáciu opísaného riadenia je potrebné toto riadenie diskretizovať a upraviť tak, aby bolo vhodné pre nasadenie v digitálnom signálovom procesore (DSP). Hlavnú výzvu predstavuje meranie prúdu multiport meniča, ktorého spínacia frekvencia bola uvažovaná na úrovni 50 kHz. Pre takéto meranie sú potrebné rýchle analógovo - digitálne prevodníky, ktoré musia vo veľmi krátkom čase poskytnúť údaj o hodnote prúdu práve riadeniu implementovanému v DSP. Okrem samotného riadenia bude tiež vhodné zaoberať sa výberom alebo návrhom vhodného planárneho transformátora pre daný multiport menič. Keďže tieto transformátory vo väčšine prípadov predstavujú zákaznicke riešenia pre konkrétnu aplikáciu, je vhodné zamyslieť sa nad možnosťou návrhu takéhoto transformátora vzhľadom na potreby multiport meniča.

REFERENCIE

- [1] Adrián Marcinek, František Ďurovský. "Matematický model výkonového multiport meniča" In: Electrical Engineering and Informatics 12 : Proceedings of the Faculty of Electrical Engineering and Informatics of the Technical University of Košice. - Košice (Slovensko) : Technická univerzita v Košiciach s. 458-464 [CD-ROM]. - ISBN 978-80-553-3940-5
- [2] Y. M. Chen, Y. C. Liu and F. Y. Wu, "Multi-input DC/DC converter based on the multiwinding transformer for renewable energy applications", *IEEE Trans. Ind. Appl.*, vol. 38, no. 4, pp. 1096-1104, Aug. 2002.
- [3] H. Wu, K. Sun, R. Chen, H. Hu and Y. Xing, "Full-bridge three-port converters with wide input voltage range for renewable power systems", *IEEE Trans. Power Electron.*, vol. 27, no. 9, pp. 3965-3974, Sep. 2012.
- [4] Z. Qian, O. Abdel-Rahman, H. Al-Atrash and I. Batarseh, "Modeling and control of three-port DC/DC converter interface for satellite applications", *IEEE Trans. Power Electron.*, vol. 25, no. 3, pp. 637-649, Mar. 2010.
- [5] Q. Xu, N. Vafamand, L. Chen, T. Dragičević, L. Xie and F. Blaabjerg, "Review on Advanced Control Technologies for Bidirectional DC/DC Converters in DC Microgrids," in *IEEE Journal of Emerging and Selected Topics in Power Electronics*, vol. 9, no. 2, pp. 1205-1221, April 2021.
- [6] S. Shao et al., "Modeling and Advanced Control of Dual-Active-Bridge DC-DC Converters: A Review," in *IEEE Transactions on Power Electronics*, vol. 37, no. 2, pp. 1524-1547, Feb. 2022.

Comparison of Questions and Answers on Stack Exchange Sites of Content Management Systems

¹Daniela HUDIKOVÁ, ²Matúš SULÍR

Department of Computers and Informatics, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Slovakia
¹daniela.hudikova@student.tuke.sk, ²matus.sulir@tuke.sk

Abstract – There exist many studies analyzing the Stack Overflow website. However, studies investigating other Stack Exchange sub-sites are much less common. This work aims to compare Stack Exchange sites focused on two rival content management systems: WordPress Development and Drupal Answers. On both sites, the response time ranged from zero seconds to multiple years. We found the most frequently used tags and monitored the development of topics over the years. We addressed the issue of the most commonly used tags, and we also found out that there was no link between the number of times a question was viewed and the number of answers to a question. Finally, we found out that the proportion of unanswered questions was slightly higher for WordPress than for Drupal.

Keywords – Stack Exchange, WordPress, Drupal, questions and answers

I. INTRODUCTION

The popularity of question-and-answer websites has grown rapidly over the years. This ecosystem is very complex and is forced to constantly adapt to new technologies. Websites known as Q&A (Questions & Answers) are often used by programmers as resources because they contain a wealth of information related to their issues. In this article, we will look at a Q&A webpage called Stack Exchange that was created on the Stack Overflow platform.

Stack Overflow is one of the most popular and well-known websites of this type. The activity on this website is moderated by a community of users, and it uses a voting system to promote high-quality content. The website was created in 2008 and has gathered a wealth of information on software development, the web industry, various programming languages, and more. The members of this community are mostly programmers, whether seniors, who with their advice and experience help to answer the questions of programming juniors, or ordinary users who do not know how to help with the problem themselves.

The aim of this work is to compare two selected Stack Exchange sites focused on content management systems (CMSs), i.e., systems used for user-friendly management of blogs, company presentations, and similar websites. The sites of interest are WordPress Development (focused on the WordPress CMS) and Drupal Answers (Drupal CMS). We will pose and answer suitable research questions covering users, brands, topics, and unanswered questions on these websites.

II. RELATED WORK

Trienes and Balog [1] addressed the topic of identifying unclear questions in Q&A communities, who stated that detecting an unclear question can be considered a binary classification problem. They proposed a classification method that took into account the characteristics of similar issues and compared them with traditional text classification methods. They defined the question as unclear when it received another clarifying question. A question was defined as clear if an answer was provided without such clarifying questions.

Jani and Peddibhotla [2] tried to find out why some questions on Stack Exchange get a lot of answers, while many questions get no answers or very few of them. To address this issue, they used the NetLogo program, which implemented an agent-based model. They found that questions that received fewer answers were also less displayed, while questions with a large number of answers were displayed quite often. This work focused on Stack Exchange, i.e., the collection of websites we work with.

Barua et al. [3] in their work tried to determine the general discussion topics of developers, the trends of topics, and the patterns of interaction between topics. They designed a methodology for discovering and quantifying topics and trends on Stack Overflow. Their methodology was based on Latent Dirichlet Allocation. They found that mobile application development was growing faster than web application development, the scripting language *PHP* was becoming increasingly popular, and the Java programming language was at the forefront of programming languages.

III. DATA

In order to find answers to our research questions, we first had to find a suitable data set. Although we had several options, we eventually opted for the Stack Exchange Data Dump [4]. This is an anonymized listing of the entire content of the Stack Exchange site, created by users. The dataset contains all websites that are formatted as separate archives, which consist of XML files. Each such archive contains user posts, users, individual votes, comments, as well as the post history and post links. All this user content is protected by the CC-BY-SA 4.0 license. We used the data from 2009 to 2021.

We only needed data from Drupal and WordPress for our analysis. The data we used was from December 2021. To process them, it was necessary to write code that we wrote in the Python programming language. The development environment we used is called PyCharm by JetBrains. To process XML data in this language, we used the module *xml.etree.ElementTree* [5], which implements a simple and efficient API (Application Programming Interface) for parsing and creating XML data.

IV. RESEARCH QUESTIONS AND FINDINGS

RQ1: What is the average time for a question to receive an accepted answer on WordPress and Drupal sites?

In addition to the average time, we were also interested in other metrics, so we formulated the following sub-question: *What was the minimum, maximum time, and quartiles of the accepted response time?* In this research question, we focused on answering the question asked by the users. More specifically, we focused on the time when the answer to the question, which was marked as the accepted answer, was received. For WordPress, our average response time was 55 hours 7 minutes and 32 seconds. It was a bit more on the Drupal site. The mean time was 72 hours, 36 minutes and 56 seconds. Therefore, the average answer time on WordPress is lower than on Drupal.

The results for WordPress were quite remarkable. Since our calculations were in seconds, the minimum value was zero seconds. The maximum time to receive an accepted answer to a question was 85109 hours 15 minutes and 48 seconds, which is 9.71 years. After calculating these values, we further calculated the quartiles. The value of the first quartile Q1 was 1584.25 seconds. The value of the second quartile Q2 was 5475.50 seconds. The value of the third quartile Q3 was 33639.25 seconds.

On the Drupal website, we came across a situation similar to that of WordPress regarding the minimal amount of time. Again, the minimum time was zero. The maximum time to receive an accepted response at Drupal was 73905 hours 21 minutes 13 seconds. If we converted it to years, it would be 8.43 years.

In this case, we sorted all times from the smallest to the largest so that we could calculate the quartiles too. The first quartile Q1 had a value of 1918 seconds. The value of the second quartile Q2 was 8577 seconds. The value of Q3 was 73066 seconds.

We manually inspected a sample of cases with the zero-second time from asking the question to getting an accepted answer. In all the analyzed cases, the author of the question was the same as the author of the answer. This could be attributed to the practice of so-called reputation collectors [6]. However, this is only one of the possible explanations.

RQ2: How has the topic trend evolved over the years?

The second research question concerned trends and how they developed over time. Since the use of badges was not so popular from the beginning, we observed low numbers of their uses in the initial stages. WordPress's most widely used tag was *php* in 2009, although it was used only once. The other most used tags in the following years were *customization*, *custom-post-types*, and *plugins*. Subsequently, the tag *php* was at the top from year 2016 to 2021. The results are shown in Fig. 1.

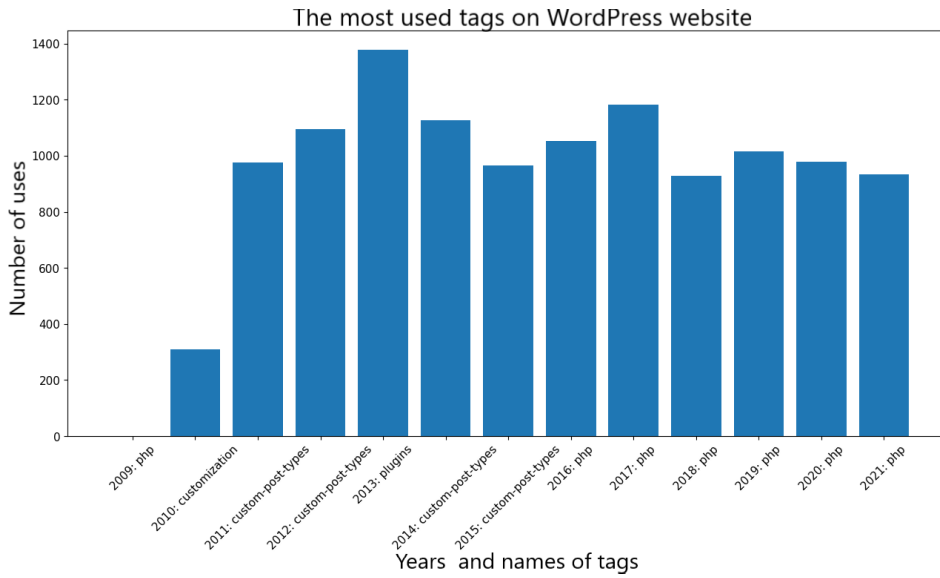


Fig. 1 The most used tags over the years on the WordPress site

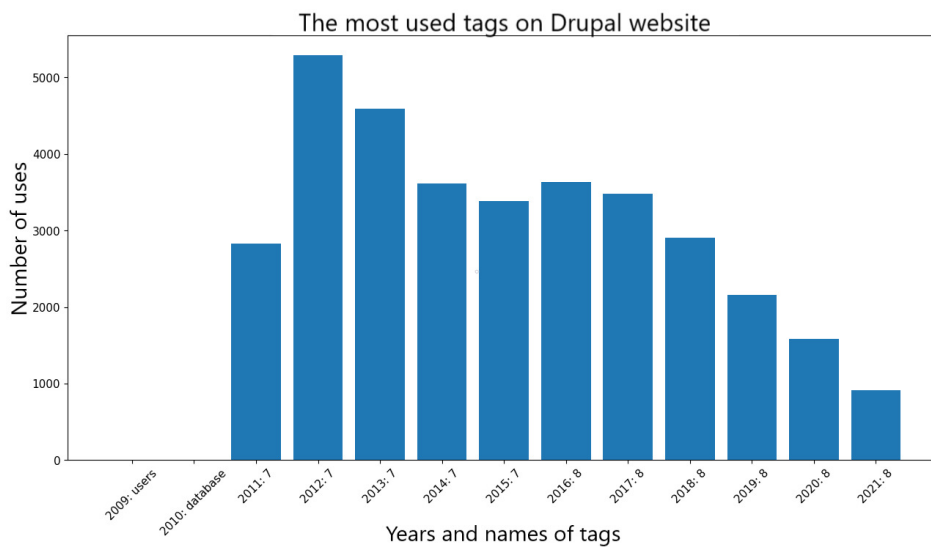


Fig. 2 The most used tags over the years on the Drupal site

The most commonly used tag for Drupal in 2009 was *users*, which was also used only once. The following year, we did not see a large increase in usage, as the most used tag was *database* with three uses, so it is clear that users were asking questions about databases. Later, only two tags were used most often: 7 (denoting the version number) and since 2016 the tag 8, due to the fact that WordPress came up with a new version that was becoming more and more popular. The results are shown in Fig. 2.

RQ3: What are the most commonly used tags on WordPress and Drupal sites?

In the third research question, we looked for the most commonly used tags on WordPress and Drupal Stack Exchange sites. This time we did not divide the brands by years, but we took into account all the years since the site has existed. The *plugins* tag ranked first on WordPress, followed by *custom-post-types*, *php*, *plugin-development*, *posts*, *functions*, *wp-query*, *customization*, *categories*, and the last in our list was the *custom-taxonomy* tag.

On the Drupal site, the “7” tag came in first, followed by the “8” tag, then *views*, *theming*, *entities*, *nodes*, *forms*, *users*, *commerce* and the last one was the *taxonomy-terms* tag.

RQ4: What are the most common topics for WordPress and Drupal?



Fig. 3 The most commonly used words in titles on the WordPress site

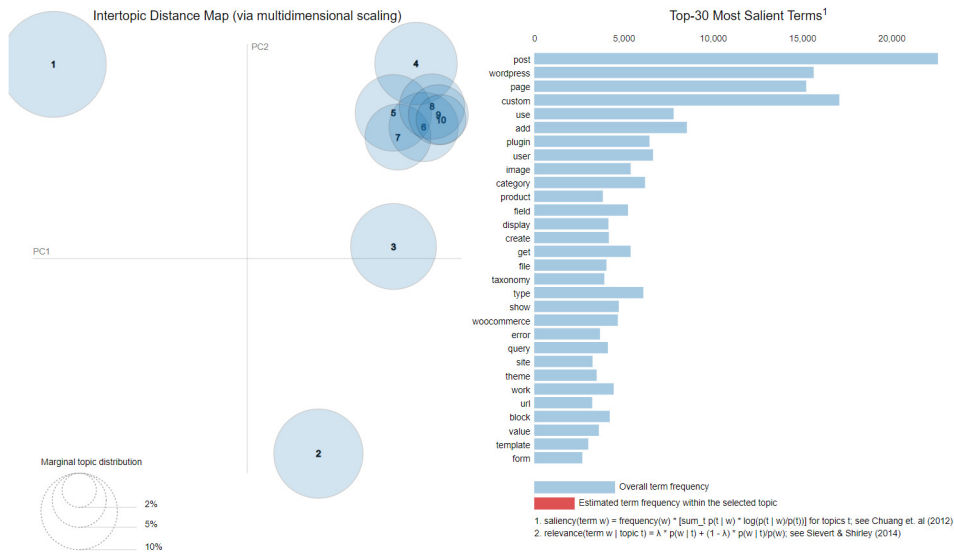


Fig. 4 WordPress topic modeling

The aim of the fourth research question was to find the most common topics. First, we extracted the title of all questions and cleaned the data by removing HTML tags and stopwords. We also applied lemmatization to obtain the base form of each word. To verify the pre-processing, we used a visual representation of our results. The most used words were displayed in the largest font. For example, on WordPress, the words *wordpress*, *custom*, *post*, *type*, *plugin*, *page*, etc. Fig. 3 shows this.

On the Drupal site, these were words like *view*, *field*, *drupal*, *node*, *module*, *add*, etc.

For this research question, we created an LDA (latent Dirichlet allocation) model with ten themes. Each topic was a combination of keywords. Each keyword contributed its weight to the topic. Later, we calculated the complexity of the model and the coherence score, which served as a measure of how good the model is. For WordPress, the complexity was -6.7448 and the topic coherence score was 0.2078 . For Drupal, the complexity was -6.8787 and the topic coherence score value was 0.2355 . The lower the value of complexity, the better the model. Although we obtained the lowest possible values when the number of topics was set to 35, we decided to set it to 10 as a practical compromise. In Fig. 4, we can see the visualization of themes for the WordPress page.

The list of the topics obtained by this process, discussed on the WordPress site, is as follows: posts, plugins, filtering, e-commerce, images, queries, forms, file displaying, database, and front-end. For Drupal, the list is the following: forms, themes, API, plugins, titles, database, content, menu, files, and classes.

RQ5: Does a high-impression question have more answers than a low-impression question?

In the fifth research question, we were interested in whether the number of question views and the number of answers are related. We wanted to find out if a question with a high number of

The relationship between the number of views and the number of answers on the WordPress website

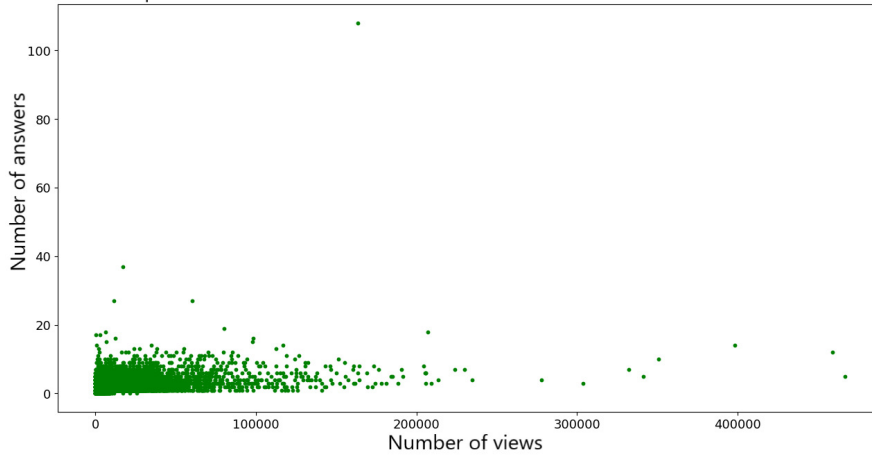


Fig. 5 A graph of the number of the answers and the number of views on the WordPress site

The relationship between the number of views and the number of ansewrs on the Drupal website

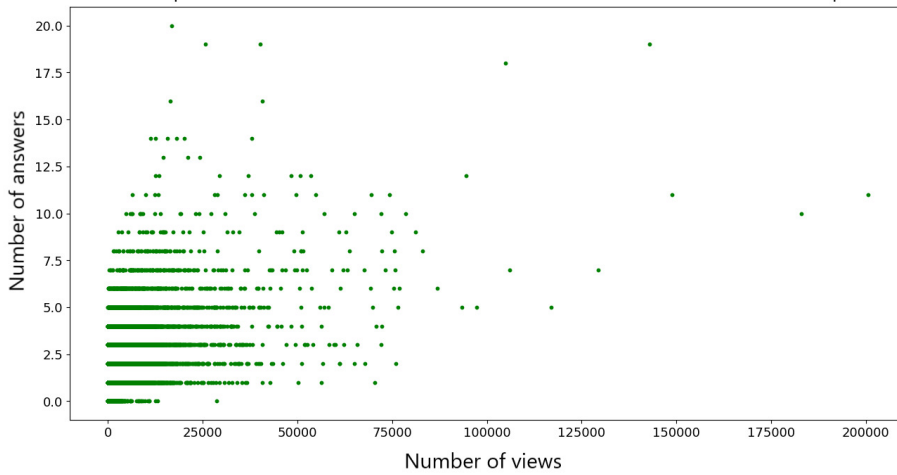


Fig. 6 A graph of the number of answers and the number of views on the Drupal site

views had more answers than a question with a low number of views. We plotted the number of views of each question on the x-axis and the number of answers on the y-axis. The results can be seen in Fig. 5 and Fig. 6. Note that overall, view counts were lower on the Drupal site than on WordPress.

To answer the research question, we calculated the correlation between the number of impressions and the number of responses. The data were not linear, and therefore we had to use the Kendall correlation coefficient. On WordPress, the correlation was equal to 0.37376 , and since this number was approaching zero, it is considered a low correlation. This result shows that our assumption was not correct, so if a question has a high number of views, it does not guarantee a high number of answers.

The data from the Drupal site were not linear either, so we used the Kendall coefficient again for the calculation. In this case, the correlation value was 0.36720 . Here, too, we can say that the correlation is low, and therefore we must refute our claim again.

RQ6: How many % of unanswered questions are on WordPress and Drupal?

The aim of the last research question was to compare the number of unanswered questions on WordPress and Drupal. At first, only questions had to be selected from all the posts. Subsequently, we selected only those that had no answer. On the WordPress page, out of the total number of $110,660$ questions, there were $17,607$ unanswered questions (15.9%). On Drupal, where the total number of questions was $85,278$, there were unanswered $12,534$ questions (14.7%). We can see the results shown in Fig. 7.

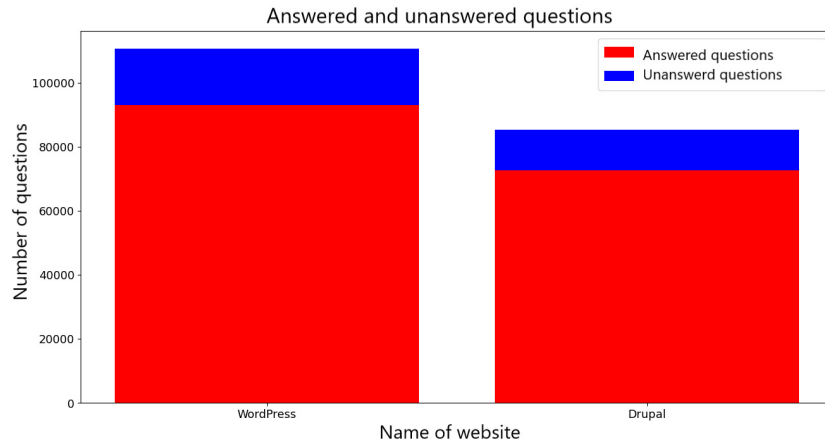


Fig. 7 A chart of answered and unanswered questions for WordPress and Drupal

V. CONCLUSION

The aim of this work was to compare two pages from the Stack Exchange platform. The first research question concerned the accepted answers on both websites. We wanted to find out what was the average time to get an accepted response, which was the minimum and maximum time, and which were the quartiles of time. We found that the average response time on WordPress was lower than on Drupal. Another research question concerned the development of trends. We searched for the most used tags and later we showed in the chart how they developed over the years. Most notably, *php* remained the most used tag on the WordPress site for six years in a row. In the third research question, we found that overall the most used tag on the WordPress site is *plugins*, while on the Drupal site the tag 7 (version number) is the most common.

The fourth research question dealt with topic modeling. We searched for the most common topics on both websites. We used an LDA-based method to address this issue. Among the topics we obtained were, for example, topics related to posts, plugins, databases, content, etc. In the next question, we looked at the number of views of the questions and their answers. We wanted to find out if a question with a high number of impressions had more answers than a question with a low number of impressions. We concluded that the two properties were unrelated and the correlation between them was low. The last research question dealt with unanswered questions. We have found that there is not much difference in the percentage of unanswered questions between the two websites, WordPress having only a slightly lower response rate.

ACKNOWLEDGMENT

This work was supported by VEGA Grant No. 1/0630/22 Lowering Programmers' Cognitive Load Using Context-Dependent Dialogs.

REFERENCES

- [1] J. Trienes and K. Balog, "Identifying unclear questions in community question answering websites," in *European Conference on Information Retrieval*. Springer, 2019, pp. 276–289.
- [2] A. Jani and N. Peddibhotla, "Why do many questions on Stack Exchange not have many answers? A preliminary agent-based model of answer contributions to Stack Exchange," in *2020 International Conference on Social Computing, Behavioral-Cultural Modeling, & Prediction and Behavior Representation in Modeling and Simulation*, 2020.
- [3] A. Barua, S. W. Thomas, and A. E. Hassan, "What are developers talking about? An analysis of topics and trends in Stack Overflow," *Empirical Software Engineering*, vol. 19, no. 3, pp. 619–654, 2014.
- [4] "Files for stackexchange," 2021, visited: 2022-01-10. [Online]. Available: <https://archive.org/download/stackexchange>
- [5] "xml.etree.elementtree — the elementtree xml api," 2022, visited: 2022-01-10. [Online]. Available: <https://docs.python.org/3/library/xml.etree.elementtree.html>
- [6] P. K. Roy, J. P. Singh, A. M. Baabdullah, H. Kizgin, and N. P. Rana, "Identifying reputation collectors in community question answering (CQA) sites: Exploring the dark side of social media," *International Journal of Information Management*, vol. 42, pp. 25–35, 2018.

Encoding Connect-4 using Quantified Boolean Formulae Revisited

¹Ivan BANK-BEDEI, ²Marián MACH

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

¹ivan.bank-bedei@student.tuke.sk, ²marian.mach@tuke.sk

Abstract — The focus of the paper is the usage of quantified Boolean formulas for creating logical models. Particularly, focus is on modelling an antagonistic turn-taking game of two players with perfect information - Connect-4. The paper deals with the model first presented by Gent and Rowley, which, however, is presented in the original source in an invalid form. Therefore, corrections of some groups of clauses enabling to achieve the validity of the given model as a whole are presented. As part of the experimentation, several experiments mentioned in the original source were performed to verify the validity and performance of the presented modifications of the model clauses.

Keywords — connect-4, logic modelling, quantified Boolean formulas, propositional logic

I. INTRODUCTION

In the field of logical modelling, propositional logic is often used, enabling to represent problem-specific conditions and constraints that must be met by a created model. This model generally has the form of a set of logical symbols (variables), which can be interpreted in two ways - either satisfied (true) or unsatisfied (false). When searching for a satisfiable model, the existential quantification of logical symbols is implicitly used – a solver searches whether there are such values of these symbols for which the given model as a whole can be interpreted as being true.

QBF (Quantified Boolean Formulae) logic is a natural extension of propositional logic. It extends propositional logic by adding two quantifiers - universal (\forall) and existential (\exists) quantifiers [1]. These quantifiers allow explicit quantification of logical formulas, where formula $\exists x \Phi$ is satisfiable only if the formula Φ is satisfiable for at least one interpretation of the symbol x , while for satisfiability of $\forall x \Phi$ must Φ be satisfiable for every possible interpretation of the symbol x .

On one hand, such an extension makes it possible to extend the application possibilities of logical modelling (for example to such areas as formal verification, gaming or classical planning [2]), but on the other hand, it increases the complexity of created logical models. Therefore, the availability of non-trivial examples of creating logical models, illustrating the use of the tools provided by QBF to model specific tasks, is important. One of the few such examples is modelling of the game Connect-4 [3]. The aim of the created model is to find a winning strategy for the starting player, which would guarantee him a win in any moves of the opponent. The cited source provides a detailed description of individual parts of the model of this game. Unfortunately, some parts of the description are not valid. As a result, the resulting model is not satisfiable - and thus the pedagogical goal is not met, because even if readers follow exactly all instructions for creating the model, the resulting model is not valid.

The aim of this paper is to correct the controversial parts of the model so that the resulting model is valid and thus the combination of text from the original source with the corrections presented here become a valuable example for those interested in modelling using QBF tools.

II. CONNECT-4 GAME

Connect-4 (known also as “four in a row”) is an antagonistic turn-based game of two players with complete information, when the state of the game is fully accessible to each of the players. It is played on a board represented by a vertically placed grid usually containing 6 rows and 7 columns (Fig.1).

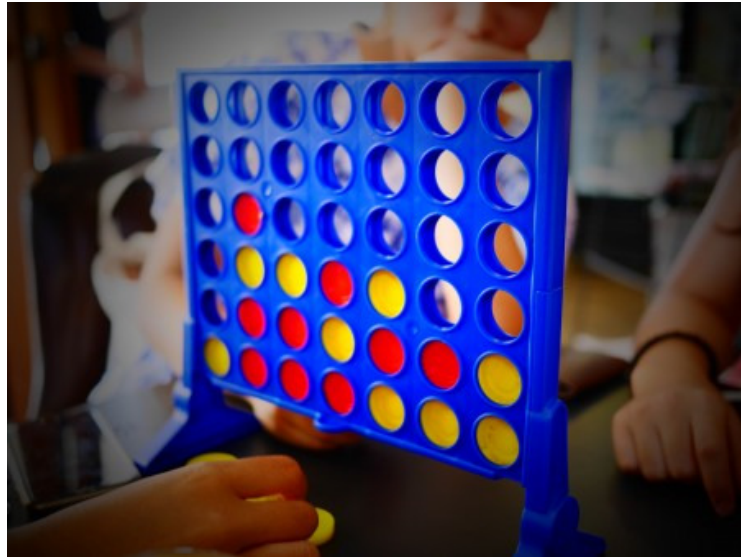


Fig. 1 Connect-4 (source: pxhere.com under Creative Commons CC0 licence).

Players alternate in their moves, with a move being represented by the insertion of one piece into one of the columns (which is not yet completely filled). The inserted piece moves itself to the lowest free position in the given column. Each player uses pieces of a separate colour. The player's goal is to achieve that after his move there is a line of four pieces of his colour, whether in the horizontal, vertical or diagonal direction. If none of the players manages to create such a line, the game ends in a draw.

For modelling purposes, a generalized version of the game is used, where

- w – number of columns
- h – number of rows
- c – the required length of a line needed to win.

The game has a maximum of $h*w$ moves (a case of a draw when the entire playing board is filled with pieces).

III. LOGICAL MODEL OF CONNECT-4

The logical model of the game Connect-4 [3] used sixteen different types of propositional symbols. Here we will list only their subset - those symbols that appear in the clauses that the source presents in an invalid form.

There were four types of symbols associated with the playing board and the pieces placed in this board. Two types of symbols characterized the position of the starting player (referred to as "red") and two types represented the position of his opposite (referred to as "white"). These symbols were *red*, *white*, *rline* and *whiteline*.

The TRUE value of a symbol $red_{x,y,z}$ / $white_{x,y,z}$ means that after making a move z , there is a red / white piece at position x,y (where x represents the row and y is the column) that has been placed on the given position either in the move z or in one of the previous moves.

Each possible line of length c (i.e. c consecutive pieces of one colour placed vertically, horizontally or diagonally) had its own identification. Thus, there were $l_v = w(h-c+1)$ vertical lines, $l_h = h(w-c+1)$ horizontal lines and $2l_d = 2(w-c+1)(h-c+1)$ diagonal lines. In addition, the colour of the pieces that occupied one of these lines was considered as well. The TRUE value of a symbol $rline_{f,z}$ / $whiteline_{f,z}$ meant that the f -th line was occupied by red / white pieces in move z .

The players' moves were represented by two types of symbols - *redmove* and *whitemove*. The TRUE value of a symbol $redmove_{y,z}$ / $whitemove_{y,z}$ meant that the given player (red or white) placed his piece in column y in move z .

If a player did not make a valid move during his turn (for example, he did not place any piece or placed a pieces in more than one column), his action was marked as cheating. This was modelled using *redcheat* and *whitecheat* symbol types. The TRUE value of a symbol $redcheat_z$ / $whitecheat_z$ indicates an invalid move by the given player in move z .

Three types of symbols were associated with the end of the game - *redwin*, *whitewin* and *gameover*. The first two types of symbols represent the win of one of the players (and thus the loss of the other player), while the value TRUE assigned to the symbol $redwin_z$ / $whitewin_z$ has

meaning that the given player won in move z (either he managed to create a line of his pieces of the required length or his opponent made an invalid move). A player's win in any move z also meant the end of the game in that move, which was represented by the $gameover_z$ symbol.

IV. INVALID CLAUSE CORRECTION

The model presented in [3] consists of clauses divided into 20 sets, with one to sixteen clauses populating one set. In total, the model contains 88 clauses. Since the analysis of all these clauses goes far beyond the scope of this paper, we will focus only on those sets and clauses that the cited source presents in the wrong form. We will present valid clause forms not in the form of disjunctive clauses as the original source did, but in the form before the transformation into conjunctive normal form, which allows better understanding of the meaning of individual clauses.

A. Clause set 14

Clauses in this set represent the relationship between a player's win in some move and that player's win in the entire game. One part of this relationship says that if a player won the game, he had to win in some move, which can be expressed by implications

$$redwin \rightarrow \bigvee_{z=1}^{hw} redwin_z \quad (1)$$

$$whitewin \rightarrow \bigvee_{z=1}^{hw} whitewin_z \quad (2)$$

or in the form of disjunctive clauses

$$\sim redwin \mid \bigvee_{z=1}^{hw} redwin_z \quad (1')$$

$$\sim whitewin \mid \bigvee_{z=1}^{hw} whitewin_z \quad (2')$$

where “ \mid ” represents logical disjunction and “ \sim ” stands for logical negation.

The source [3] in the disjunctive clauses omitted the negation of the symbols $redwin$ and $whitewin$, the clauses do not contain negation operator.

B. Clause set 15

Clauses in this set identify for each player the occurrence of lines of the player's pieces of the required length based on the occupancy of the playing board by the pieces of that player. The represented relationship has two forms:

- if a line is not occupied by a player's pieces, then at least one position of that line is without the player's piece (free or occupied by opponent's piece),
- if a line is occupied by a player's pieces, then each position of that line contains that player's piece.

Line identification makes sense only if the game is still running and therefore not finished. Otherwise, the occupancy of positions by players' pieces does not matter. Such identification of vertical and horizontal lines of one player takes the form

$$\bigwedge_{z=1}^{hw} \bigwedge_{y=1}^w \bigwedge_{x'=1}^{h-c+1} \left(\sim gameover_z \ \& \ \sim rline_{w(x'-1)+y,z} \rightarrow \bigvee_{x=x'}^{x'+c-1} \sim red_{x,y,z} \right) \quad (3)$$

$$\bigwedge_{z=1}^{hw} \bigwedge_{y=1}^w \bigwedge_{x'=1}^{h-c+1} \bigwedge_{x=x'}^{x'+c-1} \left(\sim gameover_z \ \& \ rline_{w(x'-1)+y,z} \rightarrow red_{x,y,z} \right) \quad (4)$$

$$\bigwedge_{z=1}^{hw} \bigwedge_{x=1}^h \bigwedge_{y'=1}^{w-c+1} \left(\sim gameover_z \ \& \ \sim rline_{l_v+h(y'-1)+x,z} \rightarrow \bigvee_{y=y'}^{y'+c-1} \sim red_{x,y,z} \right) \quad (5)$$

$$\bigwedge_{z=1}^{hw} \bigwedge_{x=1}^h \bigwedge_{y'=1}^{w-c+1} \bigwedge_{y=y'}^{y'+c-1} \left(\sim gameover_z \ \& \ rline_{l_v+h(y'-1)+x,z} \rightarrow red_{x,y,z} \right) \quad (6)$$

and for the other player

$$\bigwedge_{z=1}^{hw} \bigwedge_{y=1}^w \bigwedge_{x'=1}^{h-c+1} \left(\sim gameover_z \ \& \ \sim whiteline_{w(x'-1)+y,z} \rightarrow \bigvee_{x=x'}^{x'+c-1} \sim white_{x,y,z} \right) \quad (7)$$

$$\bigwedge_{z=1}^{hw} \bigwedge_{y=1}^w \bigwedge_{x'=1}^{h-c+1} \bigwedge_{x=x'}^{x'+c-1} \left(\sim gameover_z \ \& \ whiteline_{w(x'-1)+y,z} \rightarrow white_{x,y,z} \right) \quad (8)$$

$$\bigwedge_{z=1}^{hw} \bigwedge_{x=1}^h \bigwedge_{y'=1}^{w-c+1} \left(\sim gameover_z \ \& \ \sim whiteline_{l_v+h(y'-1)+x,z} \rightarrow \bigvee_{y=y'}^{y'+c-1} \sim white_{x,y,z} \right) \quad (9)$$

$$\bigwedge_{z=1}^{hw} \bigwedge_{x=1}^h \bigwedge_{y'=1}^{w-c+1} \bigwedge_{y=y'}^{y'+c-1} \left(\sim gameover_z \ \& \ whiteline_{l_v+h(y'-1)+x,z} \rightarrow white_{x,y,z} \right) \quad (10)$$

Instead of the upper boundaries $x'+c-1$ and $y'+c-1$, the original source [3] used the upper boundaries $x'+c$ and $y'+c$, thus erroneously requesting horizontal and vertical lines to be one piece longer (but surprisingly, it used the correct length for diagonal lines).

C. Clause set 16

Clauses in this set focus on identifying a player's win on a turn. The win is determined on the basis of the playing board occupancy by the players' pieces as well as on the basis of the players' cheating (making illegal moves). Possible forms of the represented relationship are:

- the win of a player means that either the player has created at least one line of the required length from his pieces or his opponent has cheated,
- if a player did not win, he did not create any line of the required length from his pieces,
- if a player did not win, then his opponent did not cheat and thus made a valid move.

Of course, a win can be only awarded to a player if he has not cheated and thus made a valid move. And, similarly to the previous set of clauses, determining the winning only makes sense if the game is still running and therefore not over. If the game is over or the reviewed player has cheated, the creation of lines with his pieces does not matter.

The first two forms of this relationship for the player *red* are

$$\Lambda_{z=1}^{hw} \left(\sim gameover_z \ \& \ \sim redcheat_z \ \rightarrow \left(redwin_z \ \rightarrow \ whitecheat_z \mid \bigvee_{f=1}^{l_v+l_h+2l_d} rline_{f,z} \right) \right) \quad (11)$$

$$\Lambda_{z=1}^{hw} \Lambda_{f=1}^{l_v+l_h+2l_d} \left(\sim gameover_z \ \& \ \sim redcheat_z \ \rightarrow \left(\sim redwin_z \ \rightarrow \ \sim rline_{f,z} \right) \right) \quad (12)$$

and for the player *white* are

$$\Lambda_{z=1}^{hw} \left(\sim gameover_z \ \& \ \sim whitecheat_z \ \rightarrow \left(whitewin_z \ \rightarrow \ redcheat_z \mid \bigvee_{f=1}^{l_v+l_h+2l_d} whiteline_{f,z} \right) \right) \quad (13)$$

$$\Lambda_{z=1}^{hw} \Lambda_{f=1}^{l_v+l_h+2l_d} \left(\sim gameover_z \ \& \ \sim whitecheat_z \ \rightarrow \left(\sim whitewin_z \ \rightarrow \ \sim whiteline_{f,z} \right) \right) \quad (14)$$

Instead of the lower limit $f=1$, the original source [3] used the lower limit $f=0$, thus adding one more line of pieces to the possible piece lines. However, the existence of this additional line was not conditioned by any specific configuration of pieces on the playing board. Thus, it was possible at any time to declare this redundant line condition fulfilled and subsequently award the win to some player, regardless of the actual distribution of the player's pieces.

D. Clause set 20

The role of the clauses in this set is to prevent the generation and testing of symmetrical situations on the playing board. If the number of columns is even, the clauses prescribe that the placement of a piece in the right part of the board should be preceded by the placement of at least one piece in the left part of the board.

And in the case of an odd number of columns, also any number of placements in the middle column can be made. In other words, if in all previous moves pieces have been placed only in the middle column, then the player on the turn cannot place his piece in the right part of the playing board.

Of course, enforcing this only makes sense if the game is not over yet and the players are not cheating on their moves.

For an odd number of columns, the clauses have the form

$$\Lambda_{z=1}^{ceil(0.5*(h+1))} \left(\sim gameover_{2z-1} \ \& \ \sim redcheat_{2z-1} \ \rightarrow \left(\Lambda_{z'=1}^{z-1} redmove_{0.5*(w+1),2z'-1} \ \& \ \Lambda_{z'=1}^{z-1} whitemove_{0.5*(w+1),2z'} \ \rightarrow \bigvee_{y=1}^{0.5*(w+1)} redmove_{y,2z-1} \right) \right) \quad (15)$$

$$\Lambda_{z=1}^{floor(0.5*(h+1))} \left(\sim gameover_{2z} \ \& \ \sim whitecheat_{2z} \ \rightarrow \left(\Lambda_{z'=1}^{z-1} redmove_{0.5*(w+1),2z'-1} \ \& \ \Lambda_{z'=1}^{z-1} whitemove_{0.5*(w+1),2z'} \ \rightarrow \bigvee_{y=1}^{0.5*(w+1)} whitemove_{y,2z} \right) \right) \quad (16)$$

where the first clause relates to odd moves made by the starting (red) player (moves indicated with $2z-1$) and the second clause relates to even moves made by the white opponent (moves indicated with $2z$).

The original source used an incorrect move indication in the last members of the clauses. Instead of $redmove_{y,2z-1} / whitemove_{y,2z}$ used a pair of symbols $redmove_{y,z} / whitemove_{y,z}$.

V. SELECTED EXPERIMENTS WITH A LOGICAL MODEL

To verify the correctness of the above presented modifications, we performed the same experiment as in the original source [3]. There were only two factors different from the original experiment:

- We used DepQBF in the role of the QBF solver [4]
- The experiments were performed on a machine with different hardware parameters (64bit CPU i7-10510U Kaby Lake architecture clocked at 3.4GHz supplemented by 16 GB of memory)

Due to this difference, it can be expected that the results achieved will differ quantitatively, but qualitatively they should remain the same.

During the experimentation, various configurations of the problem were tested, defined by the required length of the created lines (parameter c) and the size of the playing board (width w and height h). A time limit of 10 000 seconds was set for each combination, and if it was exceeded, the attempt was terminated as unsuccessful.

A. Model validity

The first aspect of the comparison was whether the solver was able to confirm the existence of a winning strategy that would guarantee the starting player to win regardless of the opponent's moves. The achieved results are in tab. 1, where the satisfiability of the created model (SAT) means the existence of a winning strategy, while its non-existence is indicated by unsatisfiability of the model (UNSAT).

Table 1
 Model satisfiability for different game configurations. SAT value represents a satisfiable model, UNSAT stays for unsatisfiable model.

c	w	h		c	w	h	
2	2	2	SAT	3	3	3	UNSAT
2	3	3	SAT	3	4	4	SAT
2	4	4	SAT	3	5	5	SAT
2	5	5	SAT	3	6	6	SAT
2	6	6	SAT	4	4	4	UNSAT
2	7	7	SAT	4	5	5	UNSAT
2	8	8	SAT	5	5	5	UNSAT

Those configurations that were also investigated in the original source (c value limited from 2 to 4) achieved the same satisfiability value as in our experiment. Based on this, the modifications presented in this paper can be considered valid.

B. Model performance

The second aspect of the comparison was the time it took the solver to confirm the feasibility or unfeasibility of the game model. The achieved results are shown in tab. 2 (in the last column of the table).

Although the achieved time values are different than in the original source, their overall trend is the same - a relatively steep increase in time demands with increasing size of the playing board. Configurations that were no longer decidable within the time constraints were achieved very quickly. This trend is also reflected in the space requirements measured as the number of symbols and the number of clauses.

The stated trend was the same in the original source, so apart from the specific absolute values, the behaviour of the models with the modifications presented by the paper is the same as the models in the original source.

Table 2
Time and space complexity of searching for a winning strategy.

c	w	h	Number of symbols	Number of clauses	Time [s]
2	2	2	238	631	0.45
2	3	3	1117	3330	0.76
2	4	4	3456	10917	0.59
2	5	5	8374	27384	0.74
2	6	6	17314	57891	1.39
2	7	7	32043	108838	2.41
2	8	8	54652	187777	5.22
3	3	3	902	2610	0.68
3	4	4	2916	9415	1.59
3	5	5	7175	24984	72.54
3	6	6	15154	55011	8770.32
4	4	4	2468	7495	0.67
4	5	5	6175	21384	97.36
5	5	5	5374	17184	17.81

VI. CONCLUSION

Based on the performed experiments, it can be stated that we managed to correct the invalid representations of some clauses in a suitable way, which gave us the opportunity to generate valid models of the game Connect-4. On the other hand, the process of identifying invalid clauses and their correction provided us with a deeper knowledge of individual elements of the model and their interconnection, which may be an advantage in the future when creating other models from other domains based on using QBF logic.

REFERENCES

- [1] H.K. Buning and U. Bubeck, "Theory of Quantified Boolean Formulas" in Handbook of satisfiability, Biere et al. (eds.), IOT Press, Amsterdam, 2009, 735-760.
- [2] I. Shaic and J. van de Pol, "Classical planning as QBF without Grounding", in Int. Workshop on Quantified, Boolean Formulas and Beyond, TU Wien, 2021.
- [3] I.P. Gent and A.G.D. Rowley, "Encoding Connect-4 using Quantified Boolean Formulae" in *Modelling and Reformulating Constraint Satisfaction Problems: Towards Systematisation and Automation*, 2003, 78-93.
- [4] F. Lonsing and U. Egly, "DepQBF 6.0: A Search-based QBF Solver Beyond Traditional QCDCL" in *Proc. of the 26th Int. Conference on Automated Deduction*, LNCS, Spriger, 2017.

Design and Implementation of a Web Application for the CMCT&II Research Group

¹Zuzana PUGELOVÁ, ¹Ján JADLOVSKÝ

¹ Department of Cybernetics and Artificial Intelligence, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Slovak Republic

¹zuzana.pugelova@tuke.sk, ¹jan.jadlovsky@tuke.sk

Abstract — This paper mainly deals with the development of web applications for the needs of the Center of Modern Control Techniques and Industrial Informatics (CMCT&II) research group, at the Department of Cybernetics and Artificial Intelligence, Faculty of Electrical Engineering and Informatics, Technical University of Košice. The paper firstly evaluates the condition of the current websites from a functional and design point of view. The presented web application is created via WordPress content management system, which was selected based on thorough analysis. Besides stylish modern design, the web application is enhanced with carefully chosen colour palette of pages, content update, and finally increased interactivity. This paper briefly defines and evaluates state of current CMCT&II website, from UX, UI and content side and describes reasoning of selection of content management system for creation of web application. Lastly, paper also includes test of web application from UX, and UI point of view done by SEO test and questionnaire.

Keywords — CMS, Wordpress, Web application, Web design, CMCT&II

I. INTRODUCTION

The web has established its importance in our society over the duration of last few years. Nowadays, it is standard for any organization to have a website, whether is it for a company, store, or educational center. Websites are developing to be more interactive and therefore are gradually becoming web applications [1]. First impression, overall appearance, and general quality UX and UI are crucial for any website. Center for Modern Control Techniques and Industrial Informatics has also realized the importance of website's design and therefore has started upgrading its websites. CMCT&II is one of two research groups belonging to the Center for Applied Cybernetics which operates at Department of Cybernetics and Artificial Intelligence, Faculty of Electrical Engineering and Informatics, Technical University of Košice. Center currently consists of six members, who provide teaching of several subjects for the study programs Intelligent Systems and Business Informatics, manage laboratories, and educational models [2]. One of CMCT&II's many websites is its main website which is a comprehensive web portal used mainly during the teaching of subjects provided by the research group, and for presentation of research activities and laboratories, and models developed and maintained by group. In this paper, we mainly focus on development of a complex web application, which is suitable for use by various scientific groups and institutions. Application is based on the WordPress content management system (CMS). The selection of WordPress was made based on a thorough analysis of the best available free CMSs. Furthermore, the paper addresses the deficiencies of previous website, design of web application via CMS WordPress and overall development of web application, including UX and UI. The paper consists of five sections. Section II focuses on the evaluation of current state of official main website of CMCT&II, whereas Section III describes the process of comparison and selection of CMS, which is best suited for powering the CMCT&II web application. Section IV is dedicated to the development of web application, including design, content, SEO and UX, testing and overall improvement of new web application. The final Section V is devoted to summarizing the entire work.

II. CURRENT STATE OF CMCT&II WEBSITE

In 2012, a website for the Center of Modern Control Techniques and Industrial Informatics was created. The website needed to meet certain standards in terms of design and functionality. Since 2012, the requirements for the look and functionality of the websites have changed. The changes made to the web have been minimal, mostly in terms of content, such as updating data of members, labs, or models [2]. Therefore, there was a need for a major update of the website.

Weaknesses from various perspectives are described as follows. From a design point of view, the original website did not meet the requirements of a modern site, whether it was elementary issues such as minimalist design, white space design, cohesive card design, user-friendly design, and others [3]. In terms of content management, the website's management is far from ideal, as it does not use a CMS. On the functional side, we can find the following shortcomings: duplicate menu which is located at the topbar and sidebar of the page, several invalid links scattered throughout different parts of the site, sidebar menu showing nested items only in the current part of the site hierarchy, and the English version of the site unmaintained for a few years.



Fig. 1 Example page of CMCT&II

The CMCT&II website consists of six parts, namely: Profile (represents the homepage), Infrastructure, Laboratories, Members, Subjects, Models, Research, CERN, Gallery, and Partners. Regarding some of the individual parts of the site, the deficiencies are as follows. In Profile, the news section is missing a paging or "View More" button to prevent the page from scrolling. The title image has a very low quality. Images in the news are not modal and are without the zoom functionality. The Infrastructure has an image of the CMCT&II infrastructure displayed in full resolution on another tab. The Subjects section is missing the non-existent overview of elective courses in individual years and semesters with their focus, and information pages with complete list of subjects needs to be updated. The Models section contains outdated data on model administrators, manufacturers, and the use of models in the pedagogical process [4].

III. SELECTION OF CONTENT MANAGEMENT SYSTEM

CMS is a software that is primarily designed to manage the content of websites, such as blogs with various specializations, information sites, e-shops, etc. [5]. The basic features of the CMS include the creation and storage of content, publishing, collaboration, and more. CMS are mostly based on databases, which are used to store information displayed on the site, generating the content of the site using scripts, or other server-side software [6]. Documents and multimedia content are stored directly in the file system that can be accessed using web requests. CMS consists of two parts, namely the presentation system, which is intended for visitors to the portal, and the administration system, which enables the management of content by logged-in users. Login is possible from any web browser, allowing the collaboration between users [7]. The group of content management systems includes several systems that have a common basis, but at the same time they clearly differ in function. CMS are divided into several groups based on system's focus. The most used type of CMS is the Web Content Management System, which supports the creation of various web pages, such as blogs, wikis, e-shops, or corporate websites [8].

To select a suitable CMS, it is necessary to determine what the requirements of the web portal for the CMCT&II research group are. These requirements are as follows: CMS must enable the profile creation of users with different access rights; management of articles with the ability to add media files and editing the date; creating many pages, subpages and articles; availability of modern responsive themes with the possibility for adjustment; tool for reorganizing multimedia

files and creating automatic and manual backups; possibility to add calendars for all center members to make individual schedules; intuitive environment suitable for people with minimal experience in managing CMS; tool for adding custom CSS and JavaScript code, tool for searching on web pages and ability to create a customized "Error 404" page.

Drupal, Joomla! and WordPress, belong to the group of the most used open-source CMS [9]. Differences between them can be found in various functions, quantity and quality of plugins, themes, and other important aspects of CMS [10].

Based on the description of each system and subsequent comparison with each other, the WordPress CMS was selected for the creation of the CMCT&II web application. Navigation and work in the administration interface is described as intuitive. Creating a site is quite straightforward as templates can be created in advance. A plugin can be used to categorize and make the content of the media library clearer [11]. WordPress outshines others CMS in qualities like documentation, number of free plugins, range of function and others [10] [13]. WordPress is also used by scientific institutions, such as the University of Washington, whose website content is like that of the CMCT&II website [12]. Another reason for choosing WordPress was that some members of the CMCT&II research group are already familiar with this CMS.

WordPress architecture: WordPress is a standard web application that is accessible to users from a web server over the Internet, using a client-server architecture. This network architecture consists of three elements that are interconnected, namely a client (browser), a server (http, application, and database server), and a computer network (ISO / OSI 7498 network model) [13]. Because the WordPress system is programmed in the PHP scripting language, the application server must support PHP of the version 4.3 and higher. That applies for WordPress version 5.7.6. The database system requires MySQL version 4.1.3 and higher. Support for the mod_rewrite module is required from the http server. The recommended servers to run WordPress are Apache and Nginx [13]. WordPress is based on a three-tier architecture, which consists of a database, application, and presentation layer. The presentation layer is what the user sees. The application layer can be considered essentially a separate WordPress, which is responsible for the correct flow of data between the browser and the database. The database layer stores all data in the MySQL database.

IV. IMPLEMENTATION OF CMCT&II WEB APPLICATION

The web application uses the latest version of WordPress 5.7. For easier editing and creation of pages, the Elementor Website Builder plugin v3.0.16 is used. From the large number of themes available for WordPress, we decided to use the Astra theme by Brainstorm Force. It is fast, lightweight, highly customizable, and the most popular free theme [14].

A. Design of the web application

Design-wise, the website has undergone a radical change. The most visible design elements include a new, complex color spectrum, whole site responsiveness, animated header, and custom scrollbar and footer. Colors are based on the logo of CMCT&II. A light shade of blue, (#CEE2EA, like Bravo Blue), and a dark shade of blue, (#021131, like Cetacean blue), were chosen for the background. Colors, such as white, black, shades of blue (#00CCFF), and red (#FB2056, #D11818), were used for text, buttons, and links. The color spectrum is shown in Fig. 2.



Fig. 2 Color spectrum of CMCT&II website

B. Content of web application

In total, the web application consists of 85 pages and 77 articles. Compared to the original website, the content of the application has changed as well. The CERN section has been added, and the redundant Partners and Gallery sections have been removed. In addition, several new entries have been added to various pages. Due to the large number of pages, we have selected only those demonstrating significant content and design changes.

The **Home site** (shown in Fig. 3) has 4 parts, which are separated from each other by a background color with an oblique gradient. The first part is a slideshow that consists of 4 hero images. The News is the second part, which consists of chronologically sorted events. The events contain a title and a date. Unlike the original website, which displays all events without any major divisions or custom pages, the new home page displays the 6 most recent events. Following 6

events can be viewed by pressing the Load More button. Each event has its own page created that can be linked to, which was not possible on the original website. The penultimate section does not have a title assigned to it but could be named as Center in Numbers. When this section comes into view for the first time, it starts the number counting animation of the number of models, labs, courses, bachelor's and master's theses and members of the center at all positions. The last section is devoted to the partner companies and organizations with which the Center collaborates.

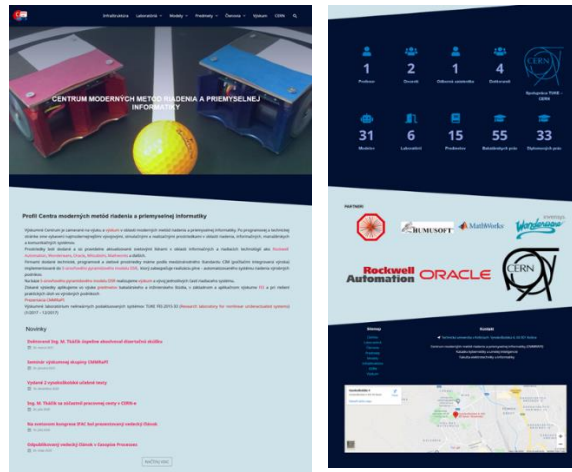


Fig. 3 Home page of CMCT&II web application

The main improvement over the original individual pages of **Laboratories and Educational laboratory models** is in the gallery, and the use of accordions that make site clearer. While the old site only had a link to the gallery on the site, we have implemented gallery with modal images, intuitive control, and other modern elements.

On the **Subjects** page there is a list of subjects which are divided into two groups, based on, whether the course is for bachelor students or magister students (Fig. 4). Below are overview tables with subjects divided into two main categories: modern control techniques and industrial informatics. Compared to the original website, overview tables have been added to the page.

		Bakalárske štúdium		Inžinierske štúdium	
	2. ročník	Modelovanie simulácia a riadenie	Príemyselná informatika	Modelovanie simulácia a riadenie	Príemyselná informatika
		Príemyselná informatika v aplikáciách		Matematika pre inteligentné systémy	Timový projekt
		Simulačné systémy		Hybridné systémy	
		Matematické metódy v priemyselnej informatike I	Senzory a sieťové členy	Simulačné systémy v mobilnej robotike	
	3. ročník	Matematické metódy v priemyselnej informatike II	Počítačové systémy	Riadenie s umelou inteligenciou	Distribučné systémy riadenia
		Príemyselná informatika a robotika			Manažérske informačné systémy
		Repetitórium	Výrobné systémy		

Fig. 4 Subjects page of CMCT&II web application

Members of the center have their own pages with basic information about them. This information includes name, phone number, email, building, room, a link to the member's ResearchGate and Pablns pages, and a photo. Additional information relating to teaching, thesis supervision or consultation, projects, and more are broken down in accordions. We see an improvement over the original site in the modern design that is visible throughout the site in the form of upgraded UX and UI, the addition of links to ResearchGate, the clear categorization of theses and publications, and the correction of incorrect and removal of non-existent links.

The **CERN** link, located in the menu of the original CMCT&II site, referred to an external site. For the new web, we didn't use the link to the external site but created a page that summarizes the goals of the collaboration, including links to other sites related to the collaboration.

The web application also includes a **search** functionality, which was not part of the original website. The functionality is positioned in the header on the right side as a magnifying glass icon. When the user clicks on icon, a box appears in which the user types the desired search term. After typing the text and pressing the enter key or by clicking the magnifying glass icon again, the search results are displayed. The functionality scans all pages of the web application.

A return HTTP code of 404 indicates that the requested page was not found. In the case of an invalid link or a link to a non-existent page within the web application, the web user is redirected to the "**Page does not exist**" page. In addition to informing the user that the link used is invalid, the page gives him the opportunity to use the search and thus try to find the desired page. This page was created using the 404page plugin.

C. Use of WordPress Plugins

For various purposes such as intuitive environment, excellent documentation, active community, tens of thousands of free extensions, WordPress plugins were used, from which a selected few are characterized. *The 404page plugin* is the most widely used plugin for creating a custom 404 error page. *The AnWPplugin* is an additional plugin for Elementor. It primarily offers the creation of post grids. *Elementor* belongs to the category of website builder plugins, i.e., plugins focused on creating pages that are intuitive, professional and with a first-class design [15]. *The iThemes Security plugin* falls into the category of security plugins and can interrupt automated attacks on the web and strengthen login security.

D. SEO test of CMCT&II web application

SEO (Search Engine Optimization) is a set of techniques that help improve the quantity of traffic to a website from search engine results such as Google or Bing [16]. SEO tester software is used to perform SEO tests. One such tester is the Seobility SEO checker, which scans websites for technical or SEO errors that could have a negative impact on search engine rankings. Seobility ranks a website based on an assigned score that reflects how well the website complies with search engines. On its official website, it states that a website with a score above 80% is highly optimized [17]. For the SEO test, we used Seobility where we achieved a score of 84% (Fig 5). To achieve this score, we needed to cut the number of loaded resources, which we achieved with Asset CleanUp: Page Speed Booster plugin. It allowed us to reduce CSS files from 22 to 10 and JS files from 29 to 12. Another tool for minimizing loading time was plugin reSmush.it Image Optimizer, which automatically performs image compression. There is the room for increasing the score by changing server from current IIS to Apache HTTP server running on Linux machine.

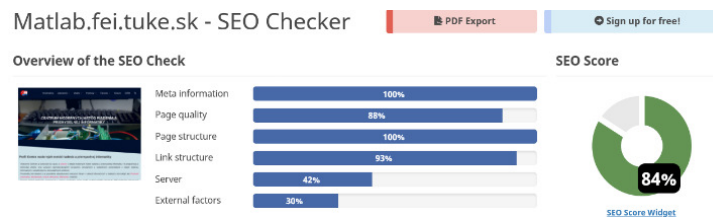


Fig. 5 Results of SEO test of CMCT&II

E. UX test of CMCT&II web application

The term UX although commonly confused with terms such as "user interface design" or "applicability", represents the overall user experience gained from using the applications. The goal of proper UX design is to navigate users between different pages easily and intuitively [18]. UX testing, otherwise known as usability testing, is a method of testing to determine how straightforward and user-friendly the application is. A small group of test users uncover errors and obstacles in the usability of the application. Research has shown that to detect 80% of problems with the application five test users will suffice [19].

To test the UX web application, we created a questionnaire, which was shared with several people. The questionnaire consisted of 18 questions and contained a link to our web application. The questions were simple and straightforward, mainly focused on finding certain specific information. Their aim was to find out the difficulty of accessing information.

The testing was done by 9 people with different computer experience, which gave us more varied results. After evaluating the results, we can say that our web application passed UX test,

although participants with low computer skills sometimes encountered a problem. Finding models seemed to be problematic, which could be explained by the fact that there are many of them, and they did not recognize the search functionality that would help them with this.

V. CONCLUSION

This paper presents an assessment of the current state of the CMCT&II website which is followed by the design and development of a new web application. The application was created using the WordPress CMS, which was selected based on the analysis and comparison with similar systems. The web application aims to present the research and educational activities of the research group. The research aspect of the created website is covered by the web pages of solved projects with a set of achieved results such as program outputs, scientific articles, and university textbooks. The educational aspect of the created website consists of web pages of laboratories, web pages of educational laboratory models, which are using CMCT&II members in teaching and students working on their thesis, and web pages of subjects, which are provided for study program Intelligent Systems by CMCT&II. An important part of the web application is the updated "members" website of the CMCT&II together with the current "infrastructure" of the center. The new innovated web meets the current standards of websites for various target groups of current and future students, new members of the group and potential colleagues from other universities. Additions of web application include search function, error page, clean design, modern galleries, CERN site, updated content, and others. The web application successfully passed the SEO test, with a score of 84% and result from UI test done by questionnaire were similarly positive. The look and design were constructed in accordance with modern standards and trends in the development of websites. In the future, the web application could be used to host various final theses, if their output is in the form of a web page.

ACKNOWLEDGMENT

This work has been supported by the project ALICE experiment at the CERN LHC: The study of strongly interacting matter under extreme conditions (ALICE KE FEI TU 0195 / 2021). The authors would like to thank Dr. Slávka Jadlovská for her valuable comments, and Dr. Lukáš Koska for his helpful advice on various technical issues.

REFERENCES

- [1] J. M. Wright, & J. Dietrich, "Survey of Existing Languages to Model Interactive Web Applications", in *APCCM* (pp. 113-123), 2008.
- [2] Z. Pugelová, "Design and implementation of web applications for the CMCT&II research group", Diploma Thesis, Technical University of Košice, Supervisor: J. Jadlovský, 2021, in Slovak.
- [3] The 8 Elements of Modern Web Design (And Web Design Trends to Watch), <https://blog.hubspot.com/marketing/elements-of-modern-web-design-list>. Accessed 2022-03-11.
- [4] Centrum moderných metód riadenia a priemyselnej informatiky, <http://kyb.feit.tuke.sk/laboratoria/>. Accessed 2022-03-11.
- [5] N. Sethi, N. Bhateja, S. Kaushal, J. Singh, "Launching WordPress-Blogs and Domains" in *International Journal of Innovative Research in Computer Science & Technology (IJIRCST)*, ISSN, 2347-5552, 2020.
- [6] J. Vivekavardhan, "Open Source Content Management System for Content Development: A Study on Wordpress, Joomla and Drupal." *Library Waves* 2.1 (2016): 6-14.
- [7] Content Management System, <https://www.optimizely.com/optimization-glossary/content-management-system/>. Accessed 2022-03-11.
- [8] M. Nakwaski, W. Zabierowski, "Content management system for web portal" 2010.
- [9] S. K. Patel, V. R. Rathod, and J. B. Prajapati. "Performance analysis of content management systems-joomla, drupal and wordpress." in *International Journal of Computer Applications*, 21.4 (2011): 39-43.
- [10] CMS comparison 2021: The most popular content management systems, <https://www.ionos.com/digitalguide/hosting/cms/cms-comparison-a-review-of-the-best-platforms/>. Accessed 2022-03-11.
- [11] Why Use WordPress? A Deep Dive Into 10 Good Reasons, <https://kinsta.com/blog/why-use-wordpress/>. Accessed 2022-03-11.
- [12] 30+ WordPress Statistics & Facts, <https://websitesetup.org/news/wordpress-statistics-facts/>. Accessed 2022-03-11.
- [13] T. Jonáš, "Aplikácia na správu výučby a obsahov", Diploma thesis, Technical University of Košice, 2015.
- [14] Astra – Fast, Lightweight & Customizable Free WordPress Theme - Download Now! <https://wpastra.com/>. Accessed 2022-03-11.
- [15] Elementor Website Builder –WordPress plugin, <https://sk.wordpress.org/plugins/elementor/>. Accessed 2022-03-11.
- [16] S. K. Patel, J. A. Patel, A. V. Patel. "Statistical analysis of seo for joomla, drupal and wordpress." in *International Journal of Computer Applications* 52.3 (2012): 1-5.
- [17] SEO Checker | Test your Website with the SEO Checker from Seobility, <https://www.seobility.net/en/seocheck>. Accessed 2022-03-11.
- [18] What is Usability Testing? UX (User Experience) Testing Example, <https://www.interaction-design.org/literature/topics/ux-design>. Accessed 2021-05-19. Accessed 2022-03-11.
- [19] What is User Experience (UX) Design?, <https://www.guru99.com/usability-testing-tutorial.html>. Accessed 2022-03-11.

Voxelizačný algoritmus generujúci SVDAG hierarchickú dátovú štruktúru

¹Vojtech KÁLAZI, ²Branislav MADOŠ, ³Anton BALÁŽ

^{1,2,3}Department of Computers and Informatics, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Slovak Republic

¹vojtech.kalazi@student.tuke.sk, ²branislav.mados@tuke.sk, ³anton.balaz@tuke.sk

Abstrakt – Článok sa venuje problematike reprezentácie geometrie voxelizovaných scén prostredníctvom doménovo-špecifických hierarchických dátových štruktúr ako sú oktantové stromy a orientované acyklické grafy. Predstavuje algoritmus umožňujúci z popisu umiestnenia vrcholov trojuholníka v dvojdimenzionálnom priestore vygenerovať popis tohto trojuholníka prostredníctvom 2D verzie hierarchickej dátovej štruktúry Sparse Voxel Directed Acyclic Graph. Táto dátová štruktúra je zostavovaná metódou zhora nadol a algoritmus sa vyznačuje nízkymi pamäťovými nárokmi v porovnaní s inými voxelizačnými algoritmi. V poslednej časti článku je uvedené vyhodnotenie testovania programovej implementácie algoritmu.

Kľúčové slová – kvadrantové stromy, Mortonovo poradie, oktantové stromy, orientované acyklické grafy, SVDAG, SVO

I. ÚVOD

V súčasnej dobe sa kladie stále väčší dôraz na čoraz väčšie rozlíšenie pri zobrazovaní objektov, aj keď to môže spôsobovať väčšiu výpočtovú náročnosť. V súčasnej dobe poznáme viacero spôsobov reprezentácie grafiky v trojdimenzionálnom priestore. Reprezentácia ako mračno bodov - táto reprezentácia sa používa hlavne ako vstupná, vytvára sa prevažne skenerom a je zvyčajne ďalej transformovaná. Ďalším spôsobom je reprezentácia vo forme polygonálnych povrchových modelov - táto reprezentácia sa používa predovšetkým pri tvorení grafického prostredia pre počítačové hry. Popisuje povrch modelovaných objektov rôznymi typmi polygónov. Tretím spôsobom je reprezentácia pomocou volumetrických dát, teda pomocou voxelov, čo možno považovať za 3D alternatívu pixelov. Tento spôsob nepopisuje objekty ako sieť polygónov, ale ako volumetrické dáta. Každý voxel má potom definované vlastnosti ako sú napríklad farba, priehľadnosť a pod. Algoritmy spracovávajúce voxelizovanú grafiku môžu byť, najmä pri veľkých rozlíšeniach voxelizovanej grafiky, časovo a pamäťovo náročné. Reprezentácia volumetrických dát je najčastejšie spájaná s hierarchickými dátovými štruktúrami ako sú doménovo-špecifické oktantové stromy a orientované acyklické grafy. Tieto hierarchické dátové štruktúry umožňujú kompaktnjšiu reprezentáciu informácií o scéne, ako je napríklad informácia o geometrii scény, teda o tom, ktoré voxely scény sú pasívne a ktoré sú aktívne.

II. KVADRANTOVÉ STROMY

V roku 1974 bola autormi R. Finkel a J.L.Bentley publikovaná dátová štruktúra, ktorú pomenovali Kvadrantový strom (z angl. QuadTree) [1]. Táto dátová štruktúra je často uvádzaná aj ako Q-strom (z angl. Q-tree). Špecifickou vlastnosťou tejto dátovej štruktúry je, že každý vnútorný uzol má práve štyri dcérske uzly. Kvadrantové stromy dekomponujú priestor rekurzívne na štyri kvadranty a najčastejšie sa používajú na dekompozíciu dvojdimenzionálneho priestoru. Táto dekompozícia končí uzlami, ktoré sa nazývajú listové uzly. Tie už obsahujú potrebnú informáciu o priestore. Jej povaha je závislá od oblasti využitia tejto dátovej štruktúry. Najvyššia úroveň stromu, reprezentovaná koreňovým uzlom stromu, predstavuje celý dvojdimenzionálny priestor, na ktorý je dekompozícia aplikovaná. Problematike kvadrantových stromov sa venovali napr. [2][3][4][5].

III. OKTANTOVÉ STROMY

Tak ako kvadrantové stromy slúžia na reprezentáciu 2D priestoru, tak pre 3D priestor bola vytvorená nová hierarchická dátová štruktúra pomenovaná oktantový strom. Oktantový strom je hierarchická dátová štruktúra navrhnutá tak, že každý nelistový uzol má práve osem potenciálnych dcérskych uzlov. Oktantové stromy sa vo všeobecnosti používajú na znázornenie vzťahu medzi objektami v trojdimenzionálnom prostredí a preto sa používajú hlavne v trojdimenzionálnej počítačovej grafike.

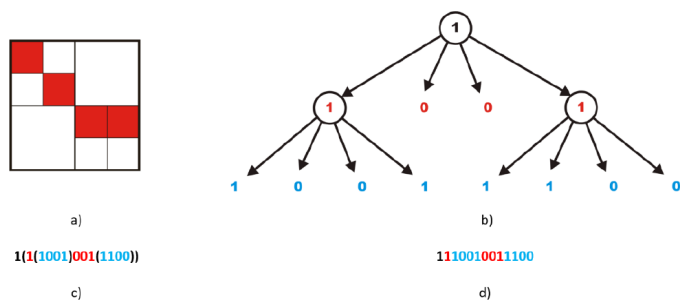


Fig. 1: Repräsentácia bez smerníkového riedkeho kvadrantového stromu voxelov [6].

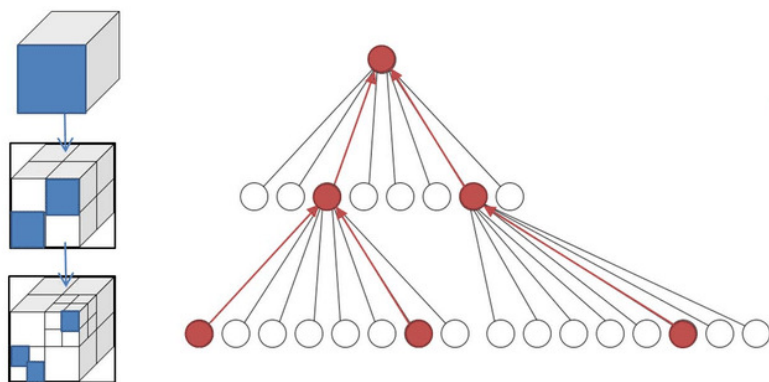


Fig. 2: Repräsentácia oktantového stromu

IV. RIEDKY ORIENTOVANÝ ACYKLICKÝ GRAF VOXELOV

Riedky orientovaný acyklický graf voxelov (z angl. Sparse Voxel Directed Acyclic Graph) je doménovo-špecifická hierarchická dátová štruktúra - orientovaný acyklický graf - ktorý vo svojej binárnej reprezentácii implementuje ukazovatele na dcérske uzly. Na rozdiel od Sparse Voxel Octrees (SVO) hierarchickej dátovej štruktúry umožňuje spájanie spoločných podstromov (common subtree merge), keď dva absolútne zhodné spoločné podstromy, sú v tejto dátovej štruktúre reprezentované tak, že plne rozvinutá reprezentácia tohto podstromu je uvedená iba raz, ako vzor a druhý podstrom je zo štruktúry vynechaný a je nahradený iba referenciou na daný vzor. Na koreň takéhoto podstromu tak môže ukazovať viac ako jeden ukazovateľ z vyššej úrovne uzlov.

V. VOXELIZAČNÝ ALGORITMUS

Popisovaný algoritmus bol navrhnutý s cieľom umožniť voxelizáciu trojuholníka a následné uloženie informácií o aktívnych voxeloch, tvoriacich daný trojuholník, do formy SVDAG hierarchickej dátovej štruktúry, tak, že nebude použitý medziprodukt, ktorým je množina voxelov, reprezentovaná vo forme ich súradníc, či už vo forme osobitne zaznamnaných súradníc pre každú dimenziu priestoru alebo vo forme Morton súradníc. Zo vstupného popisu trojuholníka tvoreného súradnicami jeho vrcholov algoritmus vytvára priamo SVDAG dátovú štruktúru, čo umožňuje v mimoriadne krátkom čase a s využitím minimálneho objemu pamäte vygenerovať SVDAG k trojuholníku a to aj v prípade mimoriadne vysokých rozlíšení, do ktorých bude voxelizovaný a to až rádo vo miliardách voxelov na hrane voxelizovanej scény.

A. Vstupné parametre

Vstupom do algoritmu sú používateľom definované tri body v dvojdimenzionálnom priestore s definovanými súradnicami x a y . Tieto tri body tvoria vrcholy trojuholníka. Ďalším vstupným parametrom je rozmer scény, ktorá má tvar štvorca. Parametrom je teda dĺžka jeho hrany v pixeloch (tá musí byť 2^n). Algoritmus bude dekomponovať scénu dovtedy, kým rozmer jedného kvadrantu nebude 2×2 . Tento rozmer je dostatočný aby sa mohli presne definovať aktívne a pasívne voxely. Ako príklad zoberieme priestor 4×4 voxelov a objekt popísaný bodmi $(0,0)$, $(0,4)$ a $(4,0)$ viď. Obr.3.

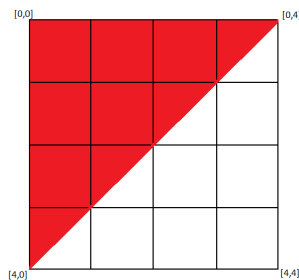
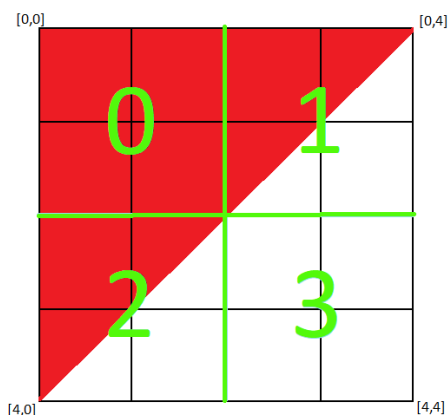


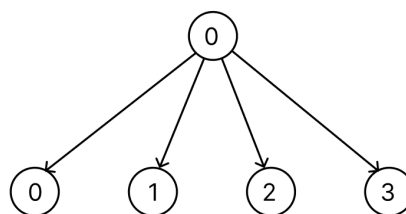
Fig. 3: Vzorový dvojdimenzionálny priestor

B. Návrh implementácie voxelizačného algoritmu

Po načítaní údajov definovaných používateľom sa obraz rozdelí na 4 kvadranty viď. Obr.4. Kvadranty budú zoradené podľa Mortonovho poradia.



(a) Rozdelenie priestoru na kvadranty



(b) Reprezentácia priestoru ako kvadrantový strom

CHNM	24b rezervovaných	Smerníky na pod kvadranty
00000111	0 24b rezervovaných	0 32bit, 1 32bit, 2 32bit, 3 32bit

(c) Binárna reprezentácia koreňového uzla ako štvorstrom

Fig. 4: Rôzne reprezentácie dekomponovaného priestoru

Rozdelením priestoru na kvadranty získame štyri rovnako veľké kvadranty a pre každý z týchto kvadrantov je potrebné vypočítať priesečníky ich ohraničení s úsečkami tvoriacimi hrany trojuholníka.

Vypočítaním priesečníkov vzniknú nové vrcholy v každom kvadrante. Na tento výpočet bola použitá metóda priesečník priamok (z angl. slope-intercept). V roku 1596 Renee Descarters [7] vo Francúzsku definoval sklon priamky. Sklon priamky uvádza, ktorým smerom priamka smeruje. Ak je hodnota sklonu negatívna, priamka smeruje dole, ak je hodnota sklonu pozitívna, tak priamka stúpa. Ak je hodnota sklonu nekonečná, tak to znamená, že priamka je rovnobežná s Y-ovou osou karteziánskeho súradnicového systému. Tieto vrcholy nemusia byť správne zoradené, preto je potrebné ich ešte pred ďalším spracovaním zoradiť v smere hodinových ručičiek. Tieto vrcholy potom tvoria vrcholy mnohouholníka umiestneného v danom kvadrante. Následne sú súradnice bodov prepočítané pre každý kvadrant tak, že každý kvadrant bude začínať v bode (0,0) obr. 5.

Ďalším krokom po vypočítaní, zoradení a transformovaní vrcholov pre každý kvadrant je zoradenie kvadrantov podľa počtu vrcholov mnohouholníkov v nich obsiahnutých, a následne zoradenie kvadrantov vzostupne podľa súradníc jednotlivých bodov mnohouholníkov v nich obsiahnutých. Kvadranty so zhodnými vrcholmi sa tak dostanú vedľa seba v poli všetkých kvadrantov. Toto pole je následne lineárne prechádzané s cieľom vyhľadať zhodné kvadranty a v ich prípade nechať v poli iba jeden ich výskyt a ostatné vynechať. Úpravou prechádzajú vždy aj smerníky, ktoré na tieto viacnásobné zhodné kvadranty ukazovali, keď po úprave ukazujú na ten kvadrant, ktorý bol v zozname ponechaný. Tým je realizované zlučovanie zhodných podstromov a súčasne je udržiavaný počet kvadrantov, ktoré je potrebné spracovávať, na nižšej úrovni, čo vedie k efektívnemu spracovávaniu tak z časového ako aj priestorového hľadiska.

Vždy po spracovaní kvadrantov jednej úrovne grafu sa proces iteratívne zopakuje vygenerovaním a zlučovaním ich dcérskych kvadrantov, ktoré sú ukladané do ďalšej vrstvy grafu, až kým v poslednej iterácii nedôjde k vygenerovaniu a zlúčeniu listových uzlov grafu.

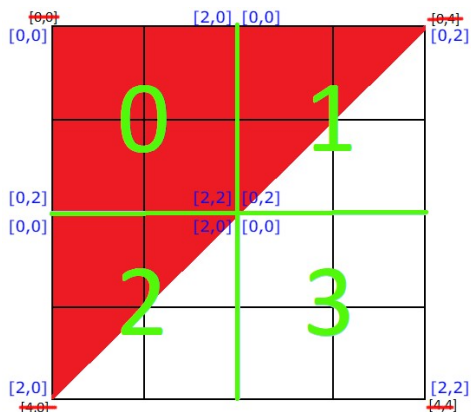


Fig. 5: Vrcholy kvadrantov po transformácii súradníc

1. 0111	0,0;2,0;2,2;0,2	0,0;0,2;2,0	0,0;2,0;0,2
2. 0111	0,0;2,0;2,2;0,2	0,0;2,0;0,2	0,0;2,0;0,2
3. 0111	0,0;2,0;2,2;0,2	0,0;2,0;0,2	

Fig. 6: Postup spracovania vrcholov

Prvý krok na Obr. 6 zobrazuje reprezentáciu masky dcérskych uzlov a vrcholy jednotlivých mnohoúhelníkov v kvadrantoch. V druhom kroku sa zoradia všetky vrcholy tak aby boli zoradené v smere hodinových ručičiek. Na obrázku je možné vidieť, že obsah kvadrantov 1 a 2 je totožný, teda môžu byť zlúčené do jedného a v ďalších krokoch bude spracovávaný iba jeden z nich. Pri zlučovaní je potrebné prepísať v rodičovskom uzle smerník ukazujúci na kvadrant číslo 2, tak aby ukazoval na kvadrant označený číslom 1.

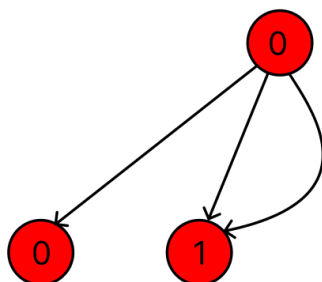


Fig. 7: Grafová reprezentácia dekomponovaného priestoru

Delenie na kvadranty prebieha rekurzívne dovtedy, kým veľkosť dekomponovaného priestoru nenadobudne rozmer 2×2 . Pri takomto rozmere priestoru je jednoduché realizovať voxelizáciu a vytvoriť tak listové uzly dátovej štruktúry. Pre uvedený príklad bolo potrebné dekompozíciu vykonať iba raz.

Výsledkom uvedeného príkladu sú dva listové uzly (Obr. 8). Jeden je obsadený vo všetkých štyroch kvadrantoch a druhý je obsadený v prvých troch kvadrantoch.

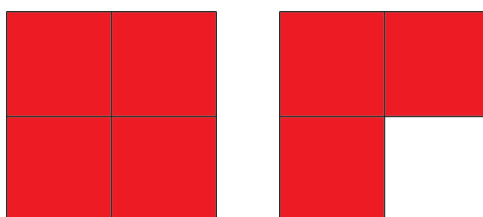


Fig. 8: Výsledné listové uzly rozmeru 2×2

C. Výstup algoritmu

Výsledok realizácie algoritmu bude reprezentovaný v binárnej reprezentácii hierarchickej dátovej štruktúry, v súboroch označených podľa úrovne od L_0 , čo určuje dekompozíciu koreňového uzla až po L_N , kde budú uložené listové uzly.

Zápis prebieha tak, že pre každý rodičovský uzol zapíšeme do dolnej polovice bajtu 4 bitovú masku dcérskych uzlov $CHNM$. Táto maska určuje, ktoré kvadranty uzla sú aktívne a ktoré pasívne.

Ak je príslušný bit nastavený na hodnotu 1, znamená to, že príslušný kvadrant je aktívny. Za maskou dcérskych uzlov nasleduje 24 prázdných bitov, ktoré sú rezervované a zaraďujú masku dcérskych uzlov na 32 bitov. V ďalšom kroku je vytvárané pole ukazovateľov na dcérske uzly. Ak je kvadrant aktívny, ukazovateľ na tento kvadrant zapíšeme do výstupného súboru v podobe 32 bitového čísla. Pri listových uzloch je uvedená $CHNM$, 24 rezervovaných bitov, ale nie je reprezentované pole ukazovateľov na dcérske uzly.

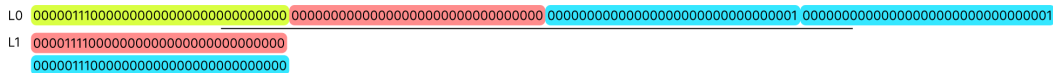


Fig. 9: Výstupné súbory pre vzorový príklad. L_0 je výstupný súbor pre koreňový uzol a v súbore L_1 sú uložené listové uzly.

VI. VYHODNOTENIE

Testovanie implementovaného algoritmu prebiehalo na rôznych veľkostiach dvojdimenzionálneho priestoru. Každý test bol spustený 12-krát, pričom bol meraný čas výpočtu, a boli odstránené najlepšie a najhoršie hodnoty. Následne zo zvyšných desiatich časov bol vypočítaný priemer a výsledná hodnota je uvedená v tabuľke nižšie.

Jedným z testovaných príkladov bol rovnoramenný pravouhlý trojuholník, ktorého pravý uhol ležal v rámci dvojdimenzionálneho priestoru v pravom dolnom rohu, ako môžeme vidieť na obr. 12.

Graduálne sme zvyšovali veľkosť daného dvojdimenzionálneho priestoru, čím sme zväčšovali tak celkový počet pixelov v danom priestore ako aj počet aktívnych pixelov tvoriacich príslušný trojuholník. Priemerné časy trvania výpočtu vo vzťahu k veľkosti scény je možné vidieť v grafe (Obr. 10). Podrobné informácie o parametroch v tomto teste sú uvedené v tabuľke na Obr. 11.

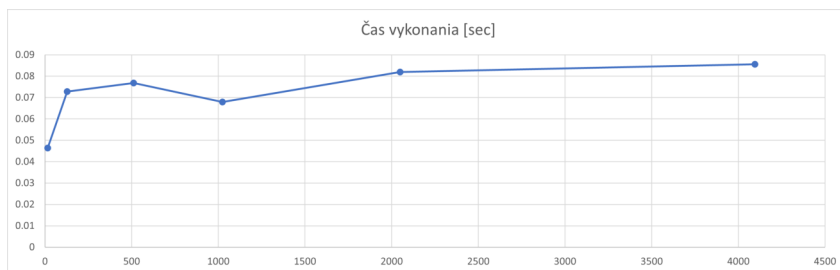


Fig. 10: Graf výsledkov

Vrcholy trojuholníku	Počet pixelov	Veľkosť hrany priestoru	Čas vykonania [sek]
(0,16),(16,0),(16,16)	256	16	0.0464
(0,128),(128,0),(128,128)	16384	128	0.0728
(0,512),(512,0),(512,512)	262144	512	0.0768
(0,1024),(1024,0),(1024,1024)	1048576	1024	0.0679
(0,2048),(2048,0),(2048,2048)	4194304	2048	0.0819
(0,4096),(4096,0),(4096,4096)	16777216	4096	0.0856

Fig. 11: Tabuľka výsledkov

VII. ZÁVER

Hlavným cieľom tejto práce bolo vytvoriť voxelizačný algoritmus, ktorý vytvára SVDAG hierarchickú dátovú štruktúru metódou zhora nadol, pričom dochádza k voxelizácii bez toho, aby ako medziprodukt vznikala veľká množina aktívnych voxelov, reprezentovaná ako súradnice týchto voxelov, ktorá by si vyžadovala veľký priestor pre jej reprezentáciu v pamäti počítača, prípadne na sekundárnom úložnom zariadení, a ktorú by bolo treba triediť.

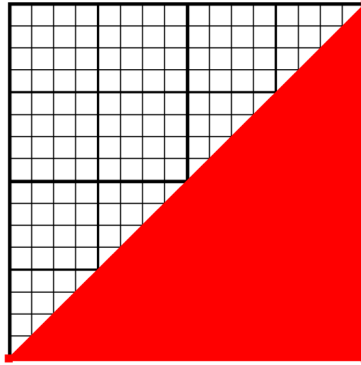


Fig. 12: Znáozornenie testovaného trojuholníka

Naopak, algoritmus prináša počas celého spracovávania trojuholníka mimoriadne úspornú reprezentáciu aktuálne rozpracovanej vrstvy generovanej SVDAG hierarchickej dátovej štruktúry, čo má pozitívny efekt aj na časovú a pamäťovú náročnosť operácií nad touto vrstvou, ktoré zahŕňajú aj proces triedenia. To otvára možnosť voxelizovať touto metódou trojuholníky v rozlíšeníach, ktoré presahujú možnosti iných algoritmov.

ACKNOWLEDGMENTS

This publication has been published with the support of the Operational Program Integrated Infrastructure within project: Research in the SANET Network and Possibilities of Its Further Use and Development (ITMS code: 313011W988), co-financed by the ERDF. This research was supported by KEGA 002TUKE-4/2021 Implementation of Modern Methods and Education Forms in the Area of Cybersecurity Towards Requirements of Labour Market.

REFERENCES

- [1] R. Finkel and J. Bentley, "Quad trees: A data structure for retrieval on composite keys." *Acta Inf.*, vol. 4, pp. 1–9, 03 1974.
- [2] R. E. Webber and M. B. Dillencourt, "Compressing quadtrees via common subtree merging," *Pattern recognition letters*, vol. 9, no. 3, pp. 193–200, 1989.
- [3] H. Samet, "The quadtree and related hierarchical data structures," *ACM Comput. Surv.*, vol. 16, no. 2, p. 187–260, 6 1984. [Online]. Available: <https://doi.org/10.1145/356924.356930>
- [4] —, "An overview of quadtrees, octrees, and related hierarchical data structures," *Theoretical Foundations of Computer Graphics and CAD*, pp. 51–68, 1988.
- [5] —, "An algorithm for converting rasters to quadtrees," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, no. 1, pp. 93–95, 1981.
- [6] L. Vokorokos, B. Madoš, and Z. Bilanová, "Psvdag: Compact voxelized representation of 3d scenes using pointerless sparse voxel directed acyclic graphs," *COMPUTING AND INFORMATICS*, vol. 39, no. 3, p. 587–616, 12 2020. [Online]. Available: http://www.cai.sk/ojs/index.php/cai/article/view/2020_3_587
- [7] D. Dennis, "René descartes curve-drawing devices: Experiments in the relations between mechanical motion and symbolic language," *Mathematics Magazine*, vol. 70, no. 3, pp. 163–174, 1997.

Robustný Prenos Dát v 5G Sieťach bez Infraštruktúry Založený na Technológii Blockchain

¹Maros BAUMGARTNER, ²Natalia KURKINA, ³Jan PAPAJ

^{1,2,3} Katedra Elektroniky a Multimediálnych Komunikácií, Fakulta Elektrotechniky a Informatiky, Technická Univerzita v Košiciach, Slovenská Republika

¹maros.baumgartenr@tuke.sk, ²natalia.kurkina@tuke.sk, ³jan.papaj@tuke.sk

Abstrakt — Mobilné Ad-Hoc siete (MANET) čelia rôznym bezpečnostným výzvam, ktoré vychádzajú zo škodlivých útokov. Ich dynamická povaha robí uzly zraniteľnejšími voči útokom zo strany škodlivých uzlov alebo vtrelcov, pretože neexistuje žiadna pevná infraštruktúra, čo vedie k tomu, že každý uzol funguje ako smerovač na prenos údajov. V súčasnosti bolo navrhnutých a implementovaných rôznymi spôsobmi niekoľko riešení zameraných na elimináciu alebo zníženie týchto škodlivých útokov. Tento dokument navrhuje model za účelom zvýšenia robustnosti dátového prenosu pomocou technológie blockchain. Keďže technológia blockchain tvorí základ nášho robustného modelu, dosiahnuté výsledky simulácií boli analyzované z pohľadu blockchainových parametrov. Dosiahnuté výsledky ukazujú, že technológia blockchain implementovaná do MANET siete zabezpečuje elimináciu škodlivých uzlov v sieti, na základe zmenených alebo zahodených dátových paketov jednotlivými uzlami. Okrem týchto parametrov sme navrhovanú sieť analyzovali aj z pohľadu počtu vygenerovaných blokov a miery škodlivých uzlov v sieti, pretože tieto parametre majú veľký vplyv na robustnosť dátovej komunikácie.

Kľúčové slová — 5G, technológia blockchain, simulácia, robustnosť

I. ÚVOD

Mobilné bezdrôtové technológie a bezdrôtová komunikácia sú v súčasnosti najdynamickejšie rozvíjajúca sa metóda slúžiaca na komunikáciu ktorú využívajú ľudia na celom svete a mimo neho. V dnešnej dobe má pri sebe takmer každý obyvateľ v civilizovanom svete zariadenie slúžiace na komunikáciu, či už to sú mobilné telefóny, laptopy, tablety, chytré hodinky a podobne. Každé z týchto zariadení je vybavené technológiami WiFi alebo Bluetooth, ktoré sú určené na komunikáciu v rámci krátkej vzdialenosti, avšak s pokrokom telekomunikačných technológií sa tieto zariadenia stávajú súčasťou mobilných bezdrôtových sietí nazývaných MANET (Mobile Ad-Hoc Network) vďaka ktorým sa vieme pripojiť k internetovému pripojeniu a komunikovať s kýmkoľvek bez ohľadu na vzájomnú vzdialenosť [1]. MANET siete disponujú určitými výhodami, ktoré sú zároveň aj ich základnými vlastnosťami. V MANET sieti môže byť každý účastník, ktorí je v tejto terminológii nazývaný uzol, zdrojom, cieľom alebo akousi prestupnou stanicou cez ktorú prúdia dátové toky od zdrojového uzla k cieľovému [2][3].

S nástupom nových typov sietí 5G a 6G sa objavili aj rôzne výzvy a oblasti výskumu, ktoré je potrebné zahrnúť pri návrhu a implementácii. Kľúčový parameter 5G a 6G sietí je decentralizácia, teda sieť, ktorá nedisponuje žiadnym centrálnym prvkom, ktorý zaoberá preposielanie dátových paketov v rámci komunikácie a, ktorý zabezpečuje plynulé a efektívne fungovanie siete. V sieťach piatej a šiestej generácie túto úlohu na seba preberajú jednotlivé uzly pripojené do siete, takýto typ komunikácie prináša mnoho výhod oproti centrálnemu prvku napríklad v rýchlosti, priepustnosti, spotrebe energie, adaptabilnosti a podobne [4]. Avšak prináša so sebou aj určité nevýhody. Jedna z najviac diskutovaných oblastí 5G a 6G sietí je vo všeobecnosti ich robustnosť a hlavne ich bezpečnosť. Z dôvodu absencie centrálného prvku, ktorý overuje všetky preposielané informácie je táto úloha prenechaná na každý uzol v sieti [5]. V takejto sieti je však omnoho jednoduchšie zaútočiť na prenášané dáta či už za účelom pozmenenia alebo vykradnutia údajov, alebo za účelom zahodenia dátového paketu kvôli snahe nedoručiť prenášané dáta od zdrojového k cieľovému uzlu [6][7].

V tomto článku sme na základe aktuálnych poznatkov o problematike odolnosti dátového prenosu v MANET sieťach v 5G a 6G sietí navrhli nový prístup, ktorý umožňuje zvýšenie

odolnosti dátového prenosu pomocou technológie blockchain. Blockchain je určitá forma databázy, ktorá sa od klasickej databázy údajov líši niekoľkými zásadnými vlastnosťami. Je to databáza, ktorá v sebe ukladá informácie o smerovaní, preposlaných paketoch, časovej stope a podobne [8]. Tieto informácie sú ukladané do blokov, ktoré tvoria vzájomnú nadväznosť čím neumožňujú pozmenenie alebo vymazanie dát. Takáto databáza blockchain je zdieľaná medzi všetkými účastníkmi siete a platí princíp, že každý vie o každom všetko. Teda ak by sa medzi uzlami objavil zlomyseľný uzol, jeho dôveryhodnosť klesne a bude vyradený z procesu komunikácie. Takýmto spôsobom sa dá jednoducho zabrániť vniknutiu akýchkoľvek narušiteľov do siete a tým výrazne zlepšiť samotnú odolnosť siete [9][10]. Tento príspevok je organizovaný nasledovne: V prvej kapitole sa venujeme súčasnej literatúre, ktorá sa zaoberá využitím technológie blockchain v rôznych viacpreskokových sieťach. V druhej kapitole popisujeme technológie MANET a blockchain, ktoré boli použité v tomto výskume. Časť tri sa zaoberá implementáciou technológie blockchain do MANET sietí a popisuje navrhnutý algoritmus tohto modelu. V poslednej časti analyzujeme a popisujeme dosiahnuté výsledky z pohľadu robustnosti dátového prenosu. Záver príspevku popisuje dosiahnuté výsledky a navrhuje ďalšie výzvy a ciele výskumu.

II. AKTUÁLNY STAV

Blockchain technológiu možno použiť aj v spojení s technológiou edge computing. Stanciu a kol. [11] implementovali systém riadenia, ktorý je distribuovaný pre edge computing s technológiou blockchain. Autori sa zamerali na štandard IEC 61499, na ktorom bol založený hierarchický distribuovaný riadiaci systém s implementovanou technológiou blockchain [12]. Autori zvolili Hyperledger Fabric ako blockchainové riešenie, v ktorom boli implementované funkčné bloky vo forme chytrých zmluv na nadradenej úrovni [13]. Následná implementácia s okrajovými uzlami pre použitie technológie edge computing, ktorá je založená na službách Docker a Kubernetes [14][15]. Pri návrhu riadiaceho systému sa zamerali na poskytovanie nízkej latencie pomocou spomínaných technológií.

Androulaki a kol. [16] predstavili rozšíriteľný modulárny blockchainový systém používaný na aplikovanie distribuovaných aplikácií. Autori implementovali číslícový zoznam na základe ktorého demonštrovali celkovú priepustnosť siete a latenciu v konkrétnych konfiguráciách. Na základe tejto technológie navrhli výpočet slúžiaci na prerušenia červích dier pomocou klastrov pre siete MANET, čím dosiahli zníženie počtu útokov červích dier na siete MANET.

V [17] Hong a kol. predstavil decentralizovaný systém určený na overenie identity s nízkou latenciou pre VEC (Vehicular Edge computing). Jednotlivé uzly v sieti (vozidlá a splnomocnené vozidlá) vypočítavajú hodnoty dôveryhodnosti decentralizovaným spôsobom, ktorý sa líši od klasickej autentifikácie. V tejto práci sa autori zamerali na sledovanie údajov o trase vozidla pomocou hash reťazca na overenie identity. Na decentralizovanú autentifikáciu bol použitý dynamický splnomocnený mechanizmus. Pomocou technológie V2V (Vehicle-to-Vehicle) uzly v sieti udržiavajú aktuálne informácie o dôveryhodnosti v blockchaine a pomocou technológie blockchain sa vyberie uzol, ktorý overí informácie o dôveryhodnosti a overovacie záznamy.

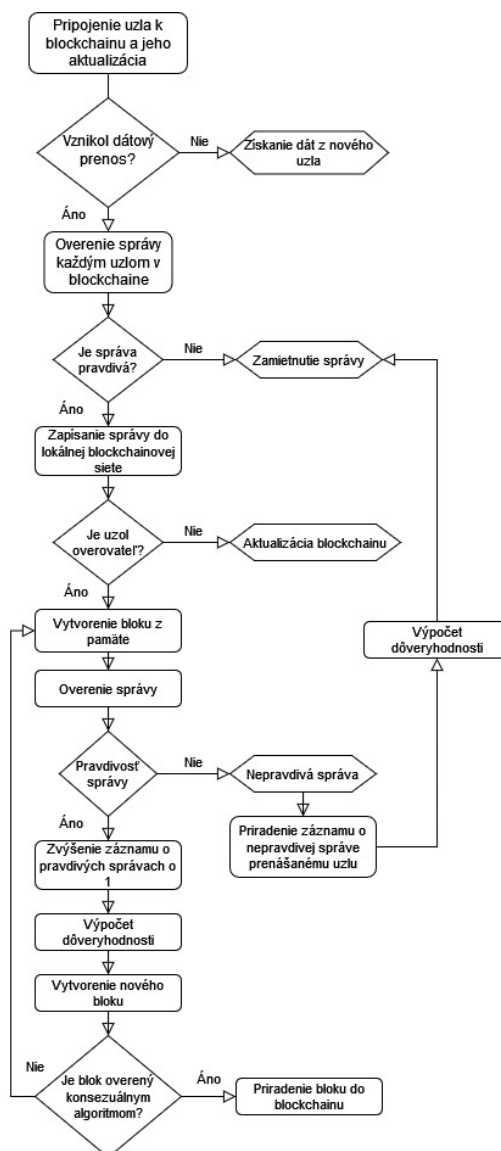
Nayana a kol. [18] predstavili vo svojej práci model založený na časových intervaloch a blockchain technológii (TIBBM – Time Interval Based Blockchain Model). Uvedený model generuje informácie, ktoré slúžia na rozlíšenie škodlivých uzlov od dôveryhodných uzlov v časových slotoch. Generovanie informácií a blokov sa spúšťa až pri komunikácii, vygenerované bloky informácií následne tvoria blockchain. Škodlivé uzly sú identifikované pomocou vybraného uzla, ktorý kontroluje správnosť blokov vytvorených uzlami v smerovacej tabuľke. Navrhovaný systém bol porovnaný so smerovaním BCR (Blockchain Contractual Routing). Na základe získaných výsledkov dosiahli výrazné zlepšenie času potrebného na vytvorenie blokov, času potrebného na identifikáciu škodlivých uzlov, zahodených paketov a presnosti detekcie škodlivých uzlov.

Shrestha a kol. [19] vo svojej práci navrhli nový systém preposielania správ v sieťach VANET. Pomocou technológie blockchain efektívne spracovávali dôveryhodnosť uzlov a dopravné informácie získané z komunikácie medzi vozidlami v sieti. Navrhovaný blockchain bol lokálne obmedzený na základe geografickej polohy. Autori sa zamerali na zníženie latencie generovania blokov v blockchaine implementáciou technológie edge computing, ktorá by umožnila využitie aplikácií v reálnom čase.

III. IMPLEMENTÁCIA TECHNOLOGIE BLOCKCHAIN DO MANET SIETI

Táto kapitola bude zameraná na implementáciu navrhnutého modelu robustného prenosu dát, ktorý pozostáva z technológie blockchain do viacpreskokovej siete. Taktiež tu bude uvedený popis simulovanej viacpreskokovej mobilnej siete a funkčných blokov technológie blockchain.

Na Obrázok 1 je znázornená navrhovaná blockchainová schéma pre MANET siete. Každý uzol, ktorý vstúpi do siete si automaticky stiahne resp. aktualizuje blockchain pretože ten funguje na princípe distribuovanej databázy medzi všetkými uzlami v sieti. Uzol, ktorý sa zúčastňuje prenosu informácií vyšle správu s konkrétnymi parametrami svojim susedným uzlom v rámci blockchainovej siete. Uzly, ktoré príjmu danú správu ju overia na základe parametrov časovej stopy, hashovacej hodnoty a paketovej hlavičky. Každý uzol zároveň overuje konkrétne prenášané informácie, aby sa zabránilo útokom na systém v podobe zaplavovania siete správami a pod. V prípade, že je správa pravdivá je takáto správa zapísaná do lokálnej blockchainovej siete. V prípade nepravdivej správy je takáto správa zamietnutá a vyradená z blockchainovej siete aby sa predišlo opätovnému overovaniu a zbytočnému zahlcovaniu siete. Pravdivá správa sa ďalej dostane k overovaciemu uzlu resp. uzlom, ktoré vytvárajú blok v pamäti a overujú správu prostredníctvom konsenzu PoW (Proof-of-Work). V prípade, že daná správa nebola overená konsenzom, je záznam o takejto správe priradený uzlu, ktorý túto správu prenášal, následne je aktualizovaná úroveň dôveryhodnosti a správa je zamietnutá. V prípade overenej správy je záznam o pravdivých resp. overených správach navýšený o 1, následne je aktualizovaná úroveň dôveryhodnosti a je vytvorený nový blok, ktorý je po úspešnom overení PoW algoritmom zapísaný do blockchainovej databázy. Na overenie tohto modelu robustného prenosu dát sme navrhli experimentálne simulácie. Cieľom týchto simulácií bolo poukázať na funkcionality, využiteľnosť a význam technológie blockchain v MANET sieťach.



Obrázok 1 Implementačný model technológie blockchain pre MANET

Predpokladáme, že sa každý uzol v sieti dokáže efektívne pripojiť k internetu a že majú dostatočný výpočtový výkon pre prípad, že by sa stali overovacími uzlami. Taktiež predpokladáme, že počet klasických uzlov preyšuje počet škodlivých uzlov, ktoré akýmkoľvek spôsobom narúšajú integritu siete, či už to je zahadzovaním dátových paketov, ich

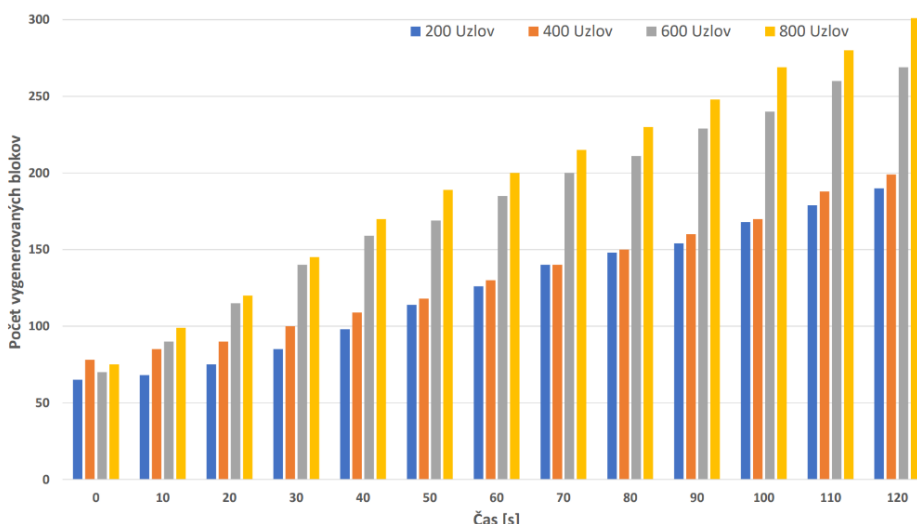
pozmeňovaním alebo prerušovaním komunikačnej trasy. Aby sa prenášaná správa považovala za správnu predpokladáme, že počet overení potrebných na potvrdenie je pätnásť. Simulačné parametre použité počas simulácií sú znázornené v Tabuľka 1 .

Premenná	Hodnota
Veľkosť oblasti [m]	2000 x 2000
Počet uzlov	200, 400, 600, 800
Rádiový dosah [m]	50
Prenosová technológia	WiFi
Trvanie simulácie [s]	120
Model mobility	Náhodný
Typ simulovanej oblasti	Ideálny (bez prekážok)
Rýchlosť pohybu uzlov [m/s]	1 – 5
Počet simulácií	100

IV. DOSIAHNUTÉ VÝSLEDKY

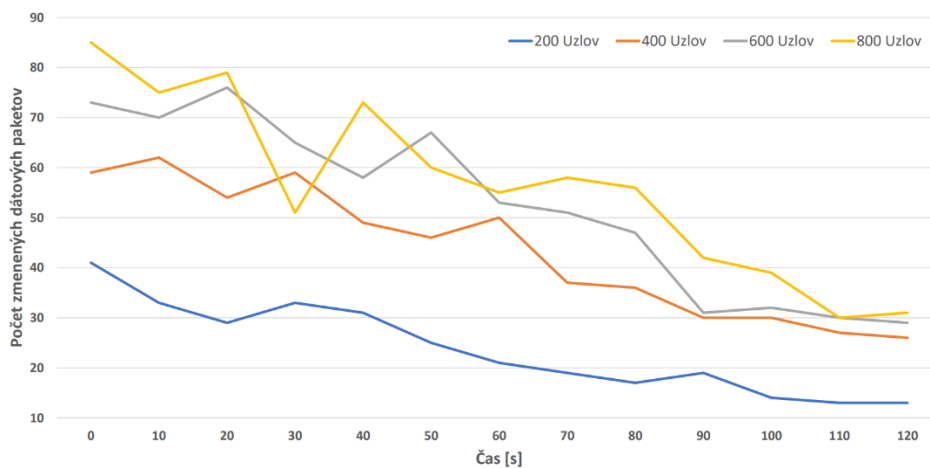
V tejto kapitole sa budeme venovať analýze dosiahnutých výsledkov prostredníctvom simulácií navrhovaného modelu umožňujúceho robustný prenos dát využitím technológie blockchain vo zvolených viacpreskokových mobilných sieťach.

Nakoľko je našim hlavným prínosom navrhnutý model robustného prenosu dát ktorého hlavná časť je tvorená technológiou blockchain rozhodli sme sa analyzovať dosiahnuté výsledky práve z pohľadu tejto technológie. Na základe toho sme analyzovali výsledky počtu vytvorených blokov blockchainu, počet zmenených a zahodených paketov počas komunikácie a počet uzlov, ktoré boli vyhodnotené ako škodlivé. Každý z týchto parametrov sme analyzovali ako pomer daného parametra k času simulácie. Tento pomer bol zvolený kvôli postupnému akcelerovaniu generovania nových blokov v blockchaine čím bola navyšovaná distribúcia tejto technológie čo viedlo k zvyšovaniu odolnosti siete.



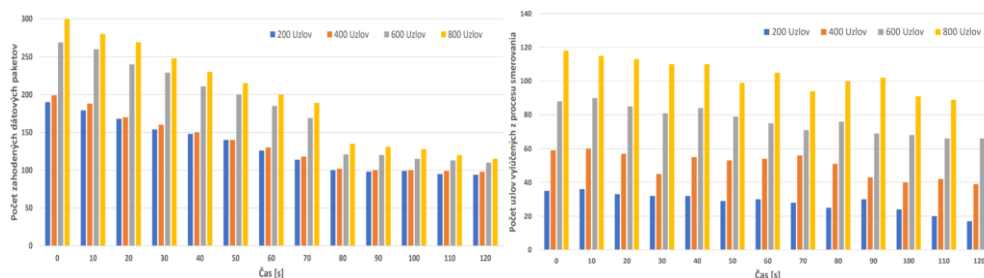
Obrázok 2 Počet vygenerovaných blokov pre všetky simulačné scenáre

Na Obrázok 2 sú znázornené dosiahnuté výsledky počtu vytvorených blokov vzhľadom k času a počtu uzlov v sieti. Na priloženom grafe môžeme vidieť, že v priebehu simulácií sa počet blokov neustále zvyšoval, miera vygenerovaných blokov závisela najmä na počte uzlov v sieti, kde sieť s 800 uzlami vygenerovala v priemere 300 blokov zatiaľ čo sieť s 200 uzlami vygenerovala v priemere 180 blokov. Táto skutočnosť bola spôsobená najmä tým, že v hustejšie obsadenej sieti bola miera vygenerovaných blokov vyššia z dôvodu väčšieho počtu overovacích uzlov, teda takáto sieť disponovala väčším výpočtovým výkonom zabezpečujúcim konsenzuálny algoritmus PoW.



Obrázok 3 Počet zmenených dátových paketov pre všetky simulačné scenáre

Technológia blockchain pozitívne vplýva aj na parameter zmenených paketov na Obrázok 3. V tomto parametri uvažujeme dátové pakety, ktoré boli vyslané zo zdrojového do cieľového uzla a ich komunikačná trasa bola zložená z jedného alebo viacerých smerovacích uzlov. V úvode simulácií bola miera zmenených paketov najvyššia, s postupným generovaním blockchainu tieto hodnoty kontinuálne klesali. Tento pozitívny jav pripisujeme práve postupnému generovaniu blokov a aktualizovaniu dôveryhodnosti jednotlivých uzlov v sieti. Z toho vyplýva, že v pokročilejších fázach simulácie pri vyššom počte úspešne resp. neúspešne prenesených paketov sa stáva sieť viac odolnou. Podobné výsledky sme zaznamenali aj pri parametri počtu zahodených paketov. Dosiahnuté výsledky počtu zahodených paketov sú zobrazené na Obrázok 4(a).



Obrázok 4 Počet zahodených dátových paketov pre všetky simulačné scenáre (a) Počet uzlov vylúčených z procesu smerovania pre všetky simulačné scenáre (b)

Parametre počtu zahodených a zmenených paketov priamo súvisia s parametrom počtu škodlivých uzlov v sieti. Čím je miera škodlivých uzlov vyššia tým je aj vyššia pravdepodobnosť ich negatívneho zásahu do komunikácie. Na Obrázok 4 (b) je znázornené zastúpenie počtu uzlov v sieti, ktoré boli vylúčené z procesu smerovania. V priebehu simulácií ich počet pozvoľne klesal čo bolo spôsobené postupným aktualizovaním ich dôveryhodnosti. Uzly, ktoré sa po opätovnej komunikácii javili ako škodlivé boli v blockchaine zapísané ako problematické a na základe týchto informácií neboli v nasledujúcich komunikáciách zvolené ako smerovacie uzly. To malo pozitívny efekt na zvýšenie robustnosti siete ako takej.

V. ZÁVER

Na základe vykonanej analýzy sme navrhli a implementovali technológiu blockchain do MANET siete. Samotná implementácia bola vykonaná prostredníctvom programovacieho jazyka Python bez použitia niektorého zo súčasne dostupných simulačných nástrojov. Na základe dosiahnutých výsledkov môžeme konštatovať, že technológia blockchain výrazným spôsobom zvyšuje robustnosť siete hlavne v oblasti bezpečnosti. Model predstavený v tomto príspevku bude slúžiť na ďalšie rozšírenie algoritmi strojového učenia za účelom ďalšieho zlepšenia robustnosti dátového prenosu.

POĎAKOVANIE

Táto publikácia vznikla vďaka podpore v rámci Operačného programu Integrovaná infraštruktúra pre projekt: Výskum v sieti SANET a možnosti jej ďalšieho využitia a rozvoja, kód ITMS: NFP313010W988, spolufinancovaný zo zdrojov Európskeho fondu regionálneho rozvoja.

REFERENCIE

- [1] P. S. R. Henrique, R. Prasad, "6G The Road to the Future Wireless Technologies 2030" River Publisher, 2021, pp. 19 – 75. ISBN:9788770224390.
- [2] K. Sheth et al. "A taxonomy of AI techniques for 6G communication networks," *Journal Computer Communications* 161, 2020.
- [3] M. Ghomsheh, A. Kamandi, "A New Metric to Evaluate Network Robustness," 6th International Conference on Web Research (ICWR). IEEE 978-1-7281-1051-6, 2020.
- [4] J. Rak, D. Hutchinson, "Guide to Disaster-Resilient Communications Networks," *Computer Communications and Networks*, Springer, 2020. ISBN 978-3-030-44685-7, pp. 47 – 141.
- [5] J. Rak, "Resilient Routing in Communication Networks," *Computer Communications and Networks*, Springer, 2020. ISBN 978-3-319-22333-9, pp. 11 – 80.
- [6] Z. Jiao, B. Zhang, L. Zhang, M. Liu, W. Gong, CH. Li, "A Blockchain Based Computing Architecture for Mobile Ad-Hoc Cloud," *IEEE Network* 0890-8044. 2020.
- [7] T.Hewa et al. "The Role of Blockchain in 6G: Challenges, Opportunities and Research Directions," *IEEE 978-1-7281-6047-4*, 2020.
- [8] Li, H. Tang, X. Wang, "Mitigating routing misbehaviour using blockchain-based distributed reputation management system for IoT networks," in: 2019 IEEE International Conference on Communications Workshops (ICC Workshops), 2019, pp. 1–6, <http://dx.doi.org/10.1109/ICCW.2019.8757083>.
- [9] C. Machado, C. M. Westphall, "Blockchain incentivized data forwarding in MANETs: Strategies and challenges," *Journal Ad Hoc Networks* 110, 2021.
- [10] David, R. Dowsley, M. Larangeira, MARS: "Monetized ad-hoc routing system (a position paper)," in: Proceedings of the 1st Workshop on Crypto currencies and Blockchains for Distributed Systems, CryBlock'18, ACM, New York, USA, 2018, pp. 82–86, <http://dx.doi.org/10.1145/3211933.3211948>.
- [11] A. Stanciu, "Blockchain based distributed control system for edge computing," in *Proc. 21st Int. Conf. Control Syst. and Comput. Sci.*, 2017, pp. 667–671.
- [12] H. Liu, Y. Zhang, and T. Yang, "Blockchain-enabled security in electric vehicles cloud and edge computing," *IEEE Netw.*, vol. 32, no. 3, pp. 78–83, May/June 2018.
- [13] L. Bach, B. Mihaljevic, and M. Zagar, "Comparative analysis of blockchain consensus algorithms," in *Proc. 41st Int. Conv. Inf. Commun. Technol., Electronics Microelectron.*, 2018, pp. 1545–1550.
- [14] J. Zou, B. Ye, L. Qu, Y. Wang, M. A. Orgun, and L. Li, "A proof-of-trust consensus protocol for enhancing accountability in crowdsourcing services," *IEEE Trans. Serv. Comput.*, vol. 12, no. 3, pp. 429–445, May/June 2018.
- [15] M. Singh and S. Kim, "Trust Bit: Reward-based intelligent vehicle communication using blockchain paper," in *Proc. IEEE 4th World Forum Internet Things*, 2018, pp. 62–67.
- [16] E. Androulaki, A. Barger, V. Bortnikov et al., "Hyperledger fabric: A distributed operating system for permissioned blockchains," in *Proceedings of the Thirteenth EuroSys Conference (EuroSys '18)*, pp. 1–15, Porto, Portugal, April 2018.
- [17] H. Liu et al., "Blockchain Empowered Cooperative Authentication With Data Traceability in Vehicular Edge Computing," *IEEE Transactions on Vehicular Technology*, 2020.
- [18] V. L. Narayana, D. Midhunchakkaravarthy, "A Time Interval based Blockchain Model for Detection of Malicious Nodes in MANET Using Network Block Monitoring Node," *International Conference on Inventive Research in Computing Applications (ICIRCA)*, 2020.
- [19] R. Shrestha et al., "A new type of blockchain for secure message exchange in VANET," *Digital Communications and Networks*, <https://doi.org/10.1016/j.dcan.2019.04.00>

Použitie štatistických testov na overenie vlastností obrazových šifrovacích algoritmov

*Jakub ORAVEC, Ľuboš OVSENÍK, Maroš LAPČÁK,
Norbert ZDRAVECKÝ, Samuel ANDREJČÍK*

Katedra elektroniky a multimediálnych telekomunikácií,
Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach

{jakub.oravec, lubos.ovsenik, maros.lapcak, norbert.zdravecky,
samuel.andrejcik}@tuke.sk

Abstrakt – Tento príspevok sa zaoberá overením možnosti použitia existujúcej skupiny štatistických testov pri vyhodnení vlastností obrazových šifrovacích algoritmov. Táto myšlienka vychádza zo skutočnosti, že v ideálnom prípade je výstupom obrazového šifrovacieho algoritmu, resp. jeho bloku pseudo-náhodná postupnosť, ktorá môže byť použitá ako vstupné dáta pre štatistické testy. Tento príspevok bude bližšie analyzovať skupinu štatistických testov NIST 800-22, ktorá bola použitá pri výbere kandidátov na šifrovací algoritmus Advanced Encryption Standard. Príspevok stručne poukáže tiež na problematiku potrebného množstva vstupných dát pri použití statických obrazov.

Kľúčové slová – databázy statických obrazov, obrazové šifrovacie algoritmy, pseudo-náhodná postupnosť, štatistické testy

I. ÚVOD

Prvé šifrovacie algoritmy vytvorené špecificky pre prácu so statickými obrazmi boli navrhnuté koncom 90. rokov minulého storočia [1]. Postupom času sa vlastnosti týchto algoritmov začali podrobne analyzovať [2], [3], čo viedlo ku zlepšeniu dosahovaných štatistických parametrov [4]. Ku týmto parametrom spočiatku patrili hlavne metriky používané pre obrazové dáta ako histogramy, korelačné diagramy, korelačné koeficienty a entropia [5]. Neskôr sa pridali aj metriky navrhnuté špeciálne pre obrazové šifrovacie algoritmy ako Number of Pixels Change Ratio (NPCR) a Unified Average Changing Intensity (UACI) [6].

V posledných rokoch sú zjavné snahy o porovnanie vlastností obrazových šifrovacích algoritmov a konvenčných riešení ako napr. Advanced Encryption Standard (AES) [7] používaných pri šifrovaní obrazových dát [8]. Z tohto dôvodu sa začali pri obrazových šifrovacích algoritmoch počítať aj hodnoty štatistických parametrov zo skupiny testov NIST 800-22 [9], ktoré boli použité pri výbere kandidátov na šifrovací algoritmus AES.

Tento príspevok v krátkosti predstaví jednotlivé testy zo skupiny NIST 800-22, predvolené hodnoty ich parametrov a spôsob ich použitia na vyhodnotenie štatistických parametrov obrazových šifrovacích algoritmov, resp. ich blokov. Okrem toho poukáže na niektoré z používaných databáz statických obrazov a objasní možnosti ich použitia. Takisto bude experimentálne overené použitie testov zo skupiny NIST 800-22 v oblasti obrazových šifrovacích algoritmov.

II. SKUPINA ŠTATISTICKÝCH TESTOV NIST 800-22

Skupina NIST 800-22 [9] obsahuje celkovo 15 štatistických testov. Testy pracujú s binárnymi dátami. Anglické názvy jednotlivých testov, ich slovenské ekvivalenty, skratky názvov používané v ďalšom texte a predvolené hodnoty parametrov jednotlivých testov sú zhrnuté v Tabuľke 1.

Testy T01 a T02 sa zameriavajú na rozloženie bitov 0 a 1 v celej postupnosti, resp. v bloku s veľkosťou M bitov. Testy T03 a T04 porovnávajú výskyt postupností zložených z rovnakých bitov, resp. najdlhšej postupnosti jednotiek s očakávanými hodnotami. Test T05 skúma lineárne závislosti medzi postupnosťami vytvorenými z celej posudzovanej postupnosti. Test T06 sa snaží odhaliť periodicitu v častiach skúmanej postupnosti. Testy T07 a T08 vyhl'adávajú v oknách s dĺžkou m bitov vopred vybrané neperiodické postupnosti. Rozdiel medzi týmito testami je v tom, či sa okná

Tabuľka 1
Názvy a predvolené hodnoty parametrov testov zo skupiny NIST 800-22

skratka	anglický názov	slovenský ekvivalent podľa [10]	hodnoty parametrov
T01	frequency (monobit)	frekvenčný test	–
T02	frequency within a block	frekvenčný test v rámci bloku	veľkosť bloku $M = 128$ bitov
T03	runs	test postupností	–
T04	longest run of ones in a block	test najdlhšej postupnosti jednotiek v bloku	–
T05	binary matrix rank	test binárnej hodnoty matíc	–
T06	discrete Fourier transform (spectral)	test diskretnou Fourierovou transformáciou	–
T07	non-overlapping template matching	test neprekrývajúcich sa vzorov	dĺžka vzoru $m = 9$ bitov
T08	overlapping template matching	test prekrývajúcich sa vzorov	dĺžka vzoru $m = 9$ bitov
T09	Maurer's universal statistic	Maurerov univerzálny štatistický test	–
T10	linear complexity	test lineárnej zložitosti	veľkosť bloku $M = 500$ bitov
T11	serial	test postupností	dĺžka postupnosti $m = 16$ bitov
T12	approximate entropy	test približnej entropie	dĺžka bloku $m = 10$ bitov
T13	cumulative sums	test kumulatívnych súčtov	–
T14	random excursions	test náhodných návštev	–
T15	random excursions variant	test variácií náhodných návštev	–

neprekrývajú (test T07) alebo prekrývajú (test T08). Test T09 vyšetruje množstvo bitov medzi rovnakými binárnymi postupnosťami, čo sa dá vnímať aj ako posúdenie možnosti kompresie. Test T10 určí minimálnu dĺžku lineárneho spätného posuvného registra, ktorý je potrebný na generovanie bloku posudzovanej postupnosti a túto minimálnu dĺžku porovná s očakávanou hodnotou. Test T11 sa sústreďuje na výskyt všetkých možných m bitových blokov, pričom vychádza z toho, že ich výskyt by mal byť približne rovnaký. Test T12 taktiež zisťuje výskyt všetkých možných m bitových blokov, ale zameriava sa na výskyt dvojíc blokov s dĺžkami m a $m + 1$. Test T13 nahradí nuly v skúmanom bloku dát za hodnotu -1 a postupne počíta súčty blokov s dĺžkami $1, 2, 3, \dots, m$. Z týchto súčtov sa vyberie maximálna hodnota a podľa nej sa rozhodne o náhodnosti celej postupnosti. Test T14 zaznamenáva počet rôznych výsledkov po súčtoch blokov z testu T13. Posledný test, T15 skúma počet výsledkov z testu T13 v prípade, ak tieto súčty prekročili určitú hranicu [9].

Testy T01 až T15 vyhodnocujú vlastnosti posudzovanej postupnosti na základe vypočítaných hodnôt, ktoré sa označujú ako P-hodnota [10] (podľa angl. P-value). Na ich výpočet sa používa značné množstvo rovníc, ktoré presahujú rozsah tohto príspevku. Rovnice s popisom ich premenných obsahuje [9].

Pre dostatočnú presnosť vypočítaných P-hodnôt odporúča [9] určité dĺžky posudzovanej postupnosti. Najväčšie nároky z tohto hľadiska majú testy T14 a T15, ktoré vyžadujú minimálne 100 čiastkových postupností, každú s dĺžkou aspoň 10^6 bitov. Pre posúdenie vlastností postupnosti je teda potrebných aspoň 10^8 bitov, čo je 12,5 megabajtu. Ak uvažujeme o postupnosti vytvorenej z nekomprimovaného obrazu, ktorý je pri obrazových šifrovacích algoritmoch bežným dátovým typom, predstavuje to obraz s rozlíšením približne 2500×1667 obrazových prvkov (op) a farebnou hĺbkou 24 bitov na op. Aj keď v súčasnosti obrazy s takýmto rozlíšením existujú, vo väčšine prípadov sú spracované algoritmi stratovej kompresie. V takto upravených obrazoch sa môžu nachádzať artefakty, ktoré negatívne vplyvajú na štatistické vlastnosti posudzovanej postupnosti.

III. DATABÁZY STATICKÝCH OBRAZOV

Pre účely testovania algoritmov spracovania obrazov existuje niekoľko známych databáz statických obrazov. Kvôli jednoduchšej distribúcii prostredníctvom internetu majú tieto databázy buď obmedzený počet obrazov, nižšie rozlíšenie obrazov alebo obrazy sú stratovo komprimované. Ako príklad možno spomenúť databázu USC-SIPI [11], ktorá sa používa na posúdenie vlastností steganografických algoritmov. Táto databáza však obsahuje iba niekoľko farebných obrazov s rozlíšením vyšším ako 512×512 op. Takisto databáza FVC2004 [12] používaná pri hodnotení vlastností biometrických algoritmov obsahuje iba obrazy v odtieňoch sivej s najvyšším rozlíšením len 640×480 op. Ďalším príkladom je databáza očných dúhoviek M. Dobeša a kol. [13], ktorá obsahuje farebné obrazy s rozlíšením 768×576 op.

Kvôli nedostatku vhodných obrazov, resp. ich databáz sa pri testovaní obrazových šifrovacích algoritmov pomocou skupiny testov NIST 800-22 pristupuje ku rôznym kompromisným rieše-

niam. Ku ním patrí spájanie viacerých obrazov do jedného s vyšším rozlíšením alebo použitie jednoducho generovaného obrazu ako napr. nulového obrazu (intenzity všetkých op sú rovné nule) s dostatočným rozlíšením [14], [15]. Takéto obrazy predstavujú pre obrazové šifrovacie algoritmy jeden z najhorších možných vstupov, keďže majú vysokú mieru korelácie medzi intenzitami op (pri nulových obrazoch je hodnota korelačného koeficientu susedných op $\rho = 1$).

IV. EXPERIMENTÁLNE POUŽITIE ŠTATISTICKÝCH TESTOV NIST 800-22

Pri našej predošlej výskumnej činnosti sme štatistické testy zo skupiny NIST 800-22 [9] využili na overenie vlastností obrazových šifrovacích algoritmov viac krát, najvýraznejšie v [14], [15]. Výsledky testov pre nulové obrazy s rozlíšením 3840×2160 op [14] a 3000×2000 op [15] a farebnou hĺbkou 24 bitov na op sú uvedené v Tabuľke 2. Nulové obrazy boli zašifrované nulovými kľúčmi s dĺžkou 128 bitov (postupnosť 128 núl). Štatistické testy používali predvolené hodnoty parametrov podľa Tabuľky 1. Pri prevode matice s intenzitami op na binárnu postupnosť sa použilo načítanie hodnôt po stĺpcoch obrazu (angl. column-major order) a zoradenie bitov s najvýznamnejším bitom vľavo (angl. big endian). Načítalo sa prvých 10^8 bitov dát, ktoré sa uložili do 100 postupností s dĺžkou po 10^6 bitov.

Tabuľka 2
Výsledky štatistických testov zo skupiny NIST 800-22 pre [14] a [15]

test	výsledok z [14]	výsledok z [15]	požadovaný výsledok
T01	99/100	99/100	96/100
T02	100/100	98/100	96/100
T03	99/100	96/100	96/100
T04	97/100	97/100	96/100
T05	99/100	98/100	96/100
T06	98/100	98/100	96/100
T07	100/100	97/100	96/100
T08	99/100	96/100	96/100
T09	97/100	98/100	96/100
T10	98/100	98/100	96/100
T11	98/100	98/100	96/100
T12	98/100	96/100	96/100
T13	99/100	99/100	96/100
T14	58/58	57/58	55/58
T15	56/58	58/58	55/58

Údaje v Tabuľke 2 poukazujú na to, že algoritmy publikované v [14] a [15] produkujú postupnosti, ktoré je možné podľa skupiny testov NIST 800-22 považovať za pseudo-náhodné. Vo všetkých prípadoch je počet postupností, ktoré vyhovejú testom dostatočný. Hraničný počet vyhovujúcich postupností nastal pri algoritme [15] a testoch T03, T08 a T12. Táto skutočnosť nastala pravdepodobne z dôvodu zjednodušenia algoritmu s cieľom jeho zrýchlenia.

Okrem počtu vyhovujúcich postupností je zaujímavé sledovať aj distribúciu P-hodnôt, najmä výskyt výrazných špičiek v postupnostiach P-hodnôt pre jednotlivé testy. To však môže byť pomerne zdĺhavé, preto sa v tomto príspevku venujeme len zhrnutiu počtu vyhovujúcich postupností.

Aj ďalšie štatistické ukazovatele ako korelačné koeficienty, entropia, NPCR a UACI ukazujú v prípade algoritmov [14], [15] na ich výhodné vlastnosti. To je z veľkej miery zapríčinené použitými pseudo-náhodnými postupnosťami, ktorých vlastnosti boli overené pomocou skupiny testov NIST 800-22.

V. ZÁVER

Tento príspevok sa zaoberal predstavením skupiny štatistických testov NIST 800-22, stručným úvodom do problematiky databáz statických obrazov a ich použitím pri vyhodnotení výsledkov obrazových šifrovacích algoritmov. Výsledky publikované v [14], [15] poukazujú na pozitívny vplyv kladne vyhodnotených postupností na vlastnosti celého obrazového šifrovacieho algoritmu. Pre získanie všeobecných záverov by však bolo potrebné súvislosť týchto skutočností preveriť na väčšej vzorke obrazových šifrovacích algoritmov, čo by však prekročilo rozsah tohto príspevku. Tejtto oblasti sa plánujeme venovať v budúcnosti.

V budúcnosti by sa náš výskumný tím chcel zamerať aj na zostavenie vlastnej databázy statických obrazov, ktoré by mohli byť experimentálne použité. Takéto obrazy by mali mať dostatočne vysoké rozlíšenie na to, aby sa dali použiť aj ako vstup pri testovaní skupinou testov NIST 800-22. To sa dá dosiahnuť napr. spracovaním RAW (z angl. surový) súborov z digitálnych

fotoaparátov a exportovaním do bezstratového súborového formátu ako napr. TIFF (z angl. Tag Image File Format). V takomto prípade by na obrazy nepriaznivo vplýval iba šum snímača fotoaparátu, jeho obmedzený dynamický rozsah a s ním súvisiaci kvantizačný šum.

POĎAKOVANIE

Táto práca bola podporená výskumným projektom FEI-2022-84 “Techniky spracovania dát vo vysokorýchlostných prenosových systémoch”.

LITERATÚRA

- [1] J. Fridrich, Symmetric Ciphers Based on Two-Dimensional Chaotic Maps, *Intl. J. of Bifurcation and Chaos*, 1998, roč. 8, č. 6, s. 1259–1284. ISSN: 0218-1274. DOI: 10.1142/S021812749800098X.
- [2] D. Arroyo, G. Alvarez, V. Fernandez, On the Inadequacy of the Logistic Map for Cryptographic Applications, v *Proc. of 10th Spanish Meeting on Cryptology and Information Security*. Salamanca (Španielsko), 2008, s. 1–6.
- [3] E. Solak, C. Çokal, O. T. Yildiz a kol., Cryptanalysis of Fridrich’s Chaotic Image Encryption, *International Journal of Bifurcation and Chaos*, 2010, roč. 20, č. 5, s. 1405-1413. ISSN: 0218-1274. DOI: 10.1142/S0218127410026563.
- [4] M. Preishuber, T. Hütter, S. Katzenbeisser a kol., Depreciating Motivation and Empirical Security Analysis of Chaos-based Image and Video Encryption, *IEEE Transactions on Information Forensics and Security*, 2018, roč. 13, č. 9, s. 2137-2150. ISSN: 1556-6013. DOI: 10.1109/TIFS.2018.2812080.
- [5] G. Chen, Y. Mao, C. K. Chui, A Symmetric Image Encryption Scheme Based on 3D Chaotic Cat Maps, *Chaos, Solitons and Fractals*, 2004, roč. 21, č. 3, s. 749-761. ISSN: 0960-0779. DOI: 10.1016/j.chaos.2003.12.022.
- [6] Y. Wu, J. P. Noonan, S. Agaian, NPCR and UACI Randomness Tests for Image Encryption, *Journal of Selected Areas in Telecommunications*, 2011, roč. 2, č. 4, s. 31-38. ISSN: 1925-2676.
- [7] FIPS 197: Advanced encryption standard. DOI: 10.6028/NIST.FIPS.197.
- [8] L. Liu, S. Miao, A New Image Encryption Algorithm Based on Logistic Chaotic Map with Varying Parameter, *SpringerPlus*, 2016, roč. 5, č. 1, s. 289-300. ISSN: 2193-1801. DOI: 10.1186/s40064-016-1959-1.
- [9] NIST SP 800-22 rev. 1a: A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications [online]. [cit. 2022-05-10]. Dostupné na internete: <https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-22r1a.pdf>.
- [10] A. Nad’, *Implementácia Diehard testov pre testovanie generátorov pseudonáhodných čísel*. Diplomová práca, Univerzita Tomáša Bati ve Zlíne, 2018, 90 s.
- [11] The USC-SIPI image database [online]. [cit. 2022-05-10]. Dostupné na internete: <http://sipi.usc.edu/database/>.
- [12] FVC2004 image database [online]. [cit. 2022-05-10]. Dostupné na internete: <http://bias.csr.unibo.it/fvc2004/databases.asp>.
- [13] M. Dobeš, J. Martinek, D. Skoupil a kol., Human Eye Localization Using the Modified Hough Transform, *Optik*, 2006, roč. 117, č. 10, s. 468-473. ISSN: 0030-4026. DOI: 10.1016/j.ijleo.2005.11.008.
- [14] J. Oravec, L. Ovseník, J. Papaj, An Image Encryption Algorithm with a Plaintext-related Quantisation Scheme, *IET Image Processing*, 2021, roč. 15, č. 9, s. 1-17. ISSN: 1751-9659. DOI: 10.1049/ipr2.12174.
- [15] J. Oravec, L. Ovseník, J. Papaj, An Image Encryption Algorithm Using Logistic Map with Plaintext-Related Parameter Values, *Entropy*, 2021, roč. 23, č. 11, s. 1373-1394. ISSN: 1099-4300. DOI: 10.3390/e23111373.

Analýza a klasifikácia behaviorálnych biometrických dát

¹Jozef BUČKO, ²Matúš PLEVA

¹ Katedra elektroniky a multimediálnych telekomunikácií, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská Republika

² Katedra elektroniky a multimediálnych telekomunikácií, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská Republika

¹jozef.bucko@student.tuke.sk, ²matus.pleva@tuke.sk

Abstrakt — V tomto článku si rozoberieme rukopis ako behaviorálnu biometrickú modalitu. Hlavný prínos je zameraný na akvizíciu biometrických údajov z rukopisu pomocou priekopníckeho zariadenia Crayonic KeyVault, ktoré má zabudovaný gyroskop, akcelerometer a tlakomer. Vďaka týmto senzorom je toto zariadenie schopné zaznamenávať pohyby rukou počas písania číslíc rukou na základe výziev samotného zariadenia. Tieto údaje v surovom formáte boli časovo ovzorkované a následne normalizované a parametrizované po oknách. Časovo ovzorkované údaje boli vizualizované a bolo pozorované, ako sa na seba podobajú dáta získané pri viacerých pokusoch písania rovnakého čísla rovnakým účastníkom a naopak ako sa líšia dáta, získané od dvoch rôznych účastníkov napriek tomu, že písali rovnaké číslo. Dáta bolo následne rozdelené na testovaciu a trénovaciu sadu. Trénovacia sada bola použitá na trénovanie modelov strojového učenia a presnosť bola overená testovacou sadou. Presnosť testovania modelov bola vyhodnotená na 91,6%.

Kľúčové slová — behaviorálna biometria, rukopis, parametrizácia, klasifikácia, rozhodovací strom, SVM, KNN, validácia

I. ÚVOD

Rapidný nárast používania digitálnych technológií so sebou prináša aj otázky bezpečnosti a ochrany údajov. V rámci informačných a komunikačných technológií existuje mnoho oblastí ochrany a bezpečnosti. Od šifrovania komunikácie po zabezpečenie prístupu pomocou autentifikačných metód, ako je heslo alebo PIN kód. Tieto autentifikačné nástroje sú založené na niečom, čo človek vie. Zvýšenie ochrany môžeme dosiahnuť pridaním ďalšieho autentifikačného prvku, ktorý môže byť založený na niečom, čo človek má, napr. karta, alebo na niečom, čím človek je. Oblasť založená na autentifikácii pomocou niečoho, čím človek je sa nazýva biometria. Tieto biometrické črty, nazývané tiež modalita môžu byť fyziologické alebo behaviorálne.

Biometrické systémy je možné aplikovať ako bezpečnostný prvok nielen v informačnom systéme. Zabezpečiť nimi môžeme napríklad aj prístup do budovy alebo prechod cez dvere. S biometrickými systémami sa môžeme stretnúť napríklad aj na letiskách a mnoho krajín sníma biometrické údaje bezprostredne pri vstupe do krajiny.

V rámci tohto článku sa budeme venovať najmä biometrii rukopisu. Rukopis patrí medzi behaviorálne biometrické modalita. Spolu s rečou patrí medzi najsilnejšie biometrické modalita zo skupiny behaviorálnych biometrických modalít.

II. RUKOPIS AKO BEHAVIORÁLNA BIOMETRICKÁ MODALITA

Identifikovať človeka v biometrii znamená rozpoznať človeka na základe jeho biometrických charakteristík. Tieto biometrické charakteristiky tiež nazývame modalita a delíme ich na fyziologické a behaviorálne. Rukopis sa radí medzi behaviorálne biometrické modalita. Ako ďalšie príklady behaviorálnych modalít by sme mohli uviesť napríklad hlas či chôdzu. Medzi fyziologické modalita radíme napríklad odtlačok prsta či dúhovku oka [1].

Rukopis je silný biometrický identifikátor a spolu s rečou patrí medzi najsilnejšie a najrozšírenejšie behaviorálne modalita. Rukopis každého jedinca je unikátny, takže môže byť použitý ako biometrický identifikátor. Rukopis je okrem identifikácie a verifikácie možné použiť tiež ako indikátor pohlavia, či je pisateľ ľavák alebo pravák či dokonca ako indikátor

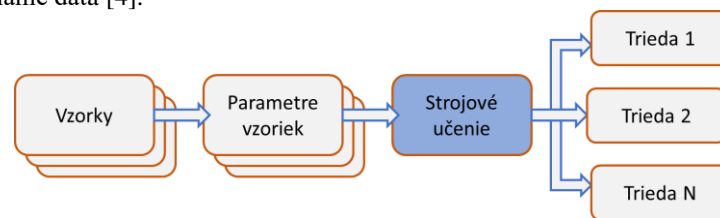
národnosti. V rukopise je tiež možné nájsť aj znaky poukazujúce na zdravotné indikácie alebo diagnózy [2].

III. BIOMETRIA A STROJOVÉ UČENIE

Biometrické systémy na identifikáciu alebo rozpoznávanie môžu byť navrhnuté a implementované za použitia techník, používajúcich algoritmy, ktoré boli manuálne naprogramované alebo pomocou techník používajúcich end-to-end tréning založený na algoritmoch strojového učenia. Pri klasických biometrických algoritmoch je správny výber čít a vlastností na klasifikáciu dôležitý pre zredukovanie variability a výpočtovej komplexnosti. Najnovšie automatické prístupy, ako strojové môžu byť výbornou alternatívou na odstránenie týchto nežiadúcich vlastností systémov [3].

Strojové učenie je procedúra tréningovania pomocou príkladov. Je to spôsob optimalizácie parametrov systému definovaných v architektúre, na splnenie kritérií za pomoci sady tréningových vzoriek. Často sa používajú štatistické techniky na to, aby mal počítač schopnosť „učiť sa“. Keď je cieľ učenia dosiahnutý, výsledný systém môže byť použitý. Takýto systém je určený na automatickú predikciu kategórie dát, jemu neznámých. Algoritmy strojového učenia sú zvyčajne delené do troch kategórií: učenie pod dohľadom, učenie bez dohľadu a posilňujúce učenie [4]. Strojové učenie je možné využiť aj v biometrii ako na Obr. 1. Vďaka algoritmom strojového učenia by sme tak mali byť schopný kategorizovať vstupné dáta do tried, pričom každá trieda bude predstavovať jeden subjekt (človeka). Na klasifikáciu dát sa využíva učenie pod dohľadom.

Pri učení pod dohľadom (Supervised learning), tréningové dáta pozostávajú zo vstupných vektorov spolu s im korešpondujúcimi cieľovými vektormi. Ak je cieľový vektor kategorizujúci, úloha sa nazýva klasifikácia alebo rozpoznávanie vzorov. Ak je cieľový vektor skutočná hodnota, úloha sa nazýva regresia. Stratové a vzdialenostné funkcie sú definované medzi aktuálnym výstupným vektorom a cieľovým vektorom pre každý vstupný vektor. Následne prebieha optimalizácia na minimalizáciu strát pre všetky tréningové vzorky. Učením systému známymi párami vstupných a cieľových vektorov od neho očakávame správnu odpoveď ak prijme neznáme dáta [4].



Obr. 1 Možné zapojenie strojového učenia v biometrii [3].

IV. AKVIZÍCIA DÁT ZARIADENÍM S IMU SENZORMI A ICH NÁSLEDNÁ NORMALIZÁCIA A PARAMETRIZÁCIA

Biometrický systém pracuje s dátami. Tieto dáta je najprv potrebné získať. Proces získavania dát sa taktiež nazýva akvizícia. Získané dáta je následne potrebné zjednotiť do jednotnej databázy, z ktorej následne môžeme extrahovať rôzne parametre vďaka ktorým môžeme dáta jednotlivých subjektov klasifikovať do jednotlivých tried [3].

Na akvizíciu dát sme použili zariadenie Crayonic KeyVault. Toto zariadenie je štandardne používané ako bezpečnostná kľúčenka na bezheslové overenie pomocou certifikátov (digitálnych podpisov) a odtlačku prsta. Zariadenie využíva štandard FIDO, ktorý samotný vznikol za cieľom redukcie závislosti verejnosti na heslách a postupne ho implementujú výrobcovia počítačov, smartfónov, internetových prehliadačov či operačných systémov [5], [6].

Zariadenie Crayonic KeyVault je v súčasnosti taktiež využívané ako experimentálne akvizíčné zariadenie zaznamenávajúce pohyby rukou pri písaní číslíc na základe výziev, ktoré dáva samotné zariadenie. Pohyby sú zaznamenávané zabudovanými senzormi ako akcelerometer, gyroskop a tlakomer. Táto meracia jednotka sa nazýva IMU jednotka. V Tabuľka 1 sú opísané veličiny ktoré pero spolu s IMU jednotkou zaznamenáva [3].

Vďaka vybaveniu pera IMU jednotkou sme tak boli schopný vykonať akvizíciu dát pohybov zaznamenaných pri písaní číslíc rukou. Tejto akvizície sa zúčastnilo 27 subjektov. Každý zo subjektov vykonal minimálne jeden pokus tréningovania, najviac však 7 pokusov. Každý pokus sa zaznamenal v samotnom CSV súbore. Výstupom jedného pokusu tréningovania tak obsahoval 9 polí. Písanie jedného čísla – jeden ťah, je časovo ovzorkovaný na 208 časových bodoch, pričom v každom časovom bode sú zaznamenané všetky veličiny. Jeden pokus tréningovania pozostával

minimálne z opakovaného písania čísl od 0 až po 9, pričom každé číslo bolo písané 5 krát. V prípade, že modul rozpoznávania čísl vyhodnotil, že ťah je chybný, mohol žiadať subjekt o opätovné písanie konkrétneho čísla. Podarilo sa nám takto získať údaje o písaní 5000 ťahov písania čísl [3].

Tabuľka 1
Veličiny a jednotky zaznamenané zariadením Crayonic KeyVault [3].

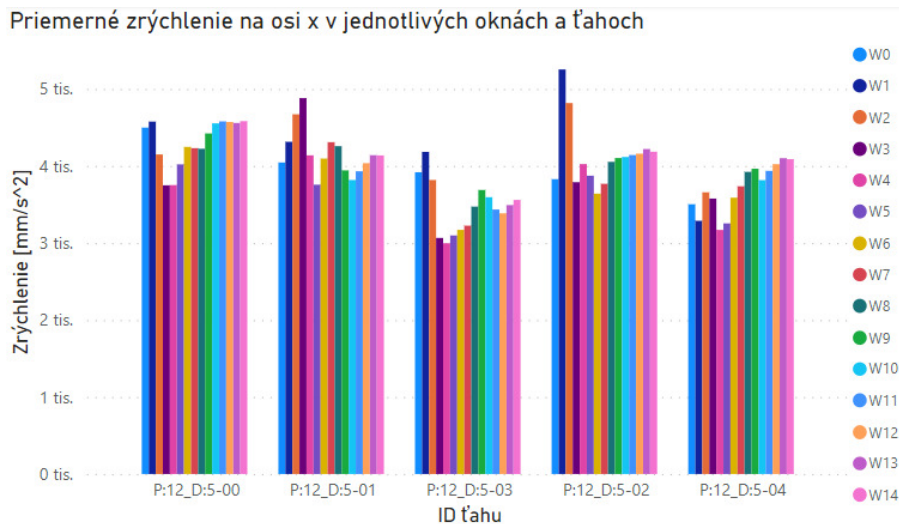
Veličina	Merná jednotka
Aktuálne písané číslo	[-]
Verzia písaného čísla	[-]
Číslo vzorky	[-]
Zrýchlenie na osiach x, y, z	[$10^{-3} \times \text{m/s}^2$]
Uhlová rýchlosť na osiach x, y, z	[$10^{-3} \times \text{rad/s}$]
Tlak hrotu na podložku	[Pa]

Údaje zo surových súborov jednotlivých pokusov tréningu boli následne zjednotené do spoločného súboru, ktorý bol obohatený o číslo pokusu tréningu a o anonymnú identitu (AID) subjektu. Obohatiť databázu o tieto indikátory bolo dôležité z hľadiska klasifikácie subjektov pomocou algoritmu strojového učenia a jednoznačné určenie konkrétneho ťahu písania [3].

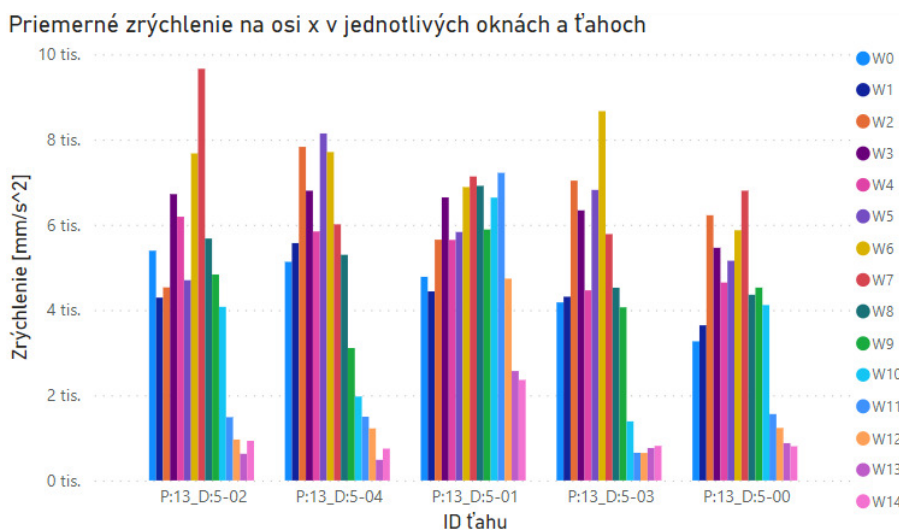
V jednom z výskumov [7] z roku 2017 bola využitá extrakcia prvkov z ručného podpisu tak, aby boli pre jedinečný podpis extrahované reprezentatívne parametre. Využili pri tom parametre podpisu ako dĺžka podpisu, výška podpisu, pomer výšky a šírky podpisu, počet súvislých ťahov v podpise, priemerný tlak, trvanie podpisu, priemerná akcelerácia a ďalšie parametre. Nami nazbierané dáta sú časovo reprezentované. Aby sme v čo najväčšej miere brali v úvahu aj časovú variabilitu ťahu pri písaní čísl, tak dáta rozdelíme do okien. Z databázy, kde je v každom riadku reprezentovaný jeden časový okamih signálu a každý stĺpec reprezentuje samostatnú veličinu (tlak, zrýchlenie na osi x, uhlová rýchlosť podľa osy z atď.) tak dostaneme databázu, v ktorej jeden riadok reprezentuje jeden ťah (napísanie jedného čísla). V stĺpcoch potom budeme mať agregované parametre, ktoré budú v každom stĺpci reprezentovať iné okno. Každý jeden ťah bol reprezentovaný 208 časovými bodmi, ktoré sme rozdelili na 16 fragmentov, pričom každý z fragmentov obsahoval 13 časových bodov. Následne sme tieto fragmenty spojili do okien, tak, že každé okno pozostáva z dvoch fragmentov, pričom každé okno (okrem prvého resp. posledného) sa prvou polovicou prekrýva s predchádzajúcim oknom a druhou polovicou s nasledujúcim oknom. Agregáciou dvoch po sebe idúcich fragmentov vzniklo jedno okno. Dve po sebe nasledujúce okná sa navzájom prekrývajú a zdieľajú jeden celý fragment. Vo výsledku je tak okno agregáciou 26 časových bodov. Každý ťah je reprezentovaný 15timi oknami, ktoré sú v databáze označené ako W0, W1 až W14. Ako dve hlavné agregačné funkcie sme využili aritmetický priemer absolútnej hodnoty a kvadratický priemer absolútnej hodnoty (ang. Root mean square) [3].

V. ANALÝZA NAZBIERANÝCH ÚDAJOV

Na Obr. 2 môžeme vidieť graf, na ktorom je znázornené priemerné zrýchlenie na osi x pri písaní čísla 5 tým istým používateľom vo viacerých ťahoch. AID ťahu je spojením AID osoby, ktorá vykonávala písanie, čísla, ktoré bolo písané a čísla pokusu. Môžeme vidieť, že pri písaní toho istého čísla tým istým používateľom sa priemerná hodnota zrýchlenia v konkrétnom okne v jednotlivých ťahoch na seba podobá. Na Obr. 3 môžeme vidieť rovnaké údaje, ktoré vznikli písaním taktiež čísla 5, len iným používateľom. Môžeme badať, že dáta z rôznych pokusov od jedného používateľa sa na seba podobajú ale dáta od rôznych používateľov, získané pri písaní rovnakého čísla sú odlišné [3].



Obr. 2 Zrýchlenie na osi x používateľa s AID 12 pri písaní číslice 5 [3].



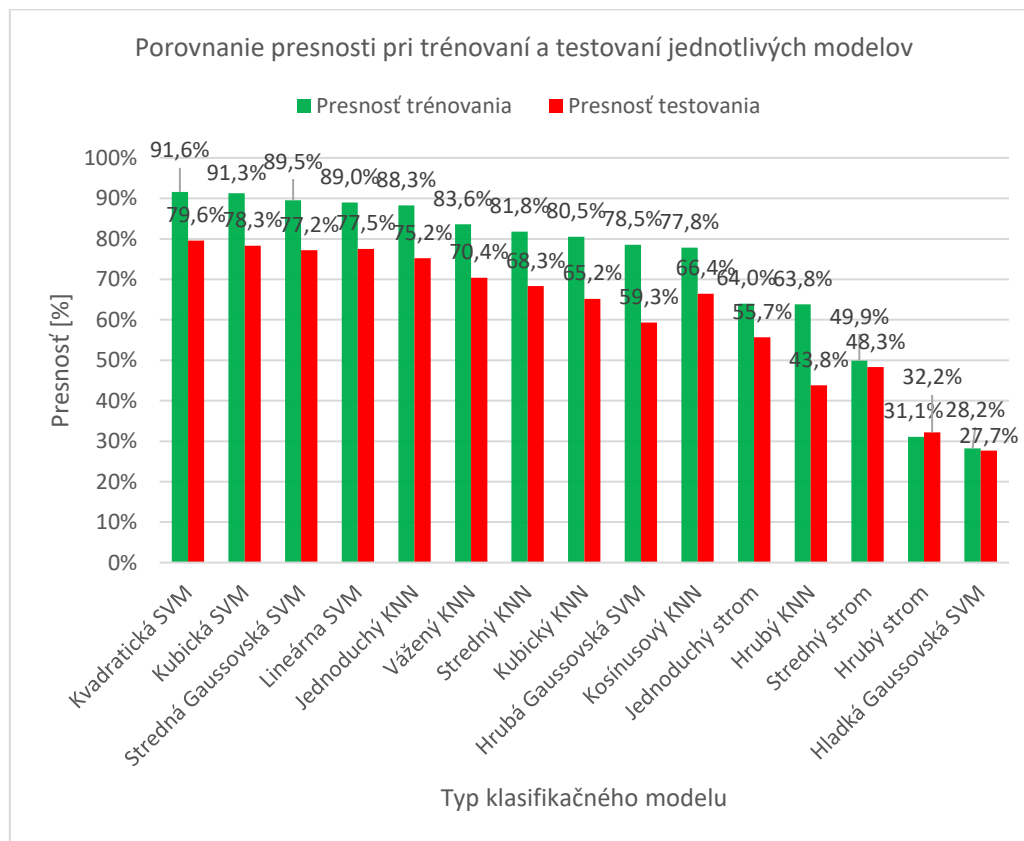
Obr. 3 Zrýchlenie na osi x používateľa s AID 13 pri písaní číslice 5 [3].

VI. KLASIFIKÁCIA ÚDAJOV ALGORITMAMI STROJOVÉHO UČENIA

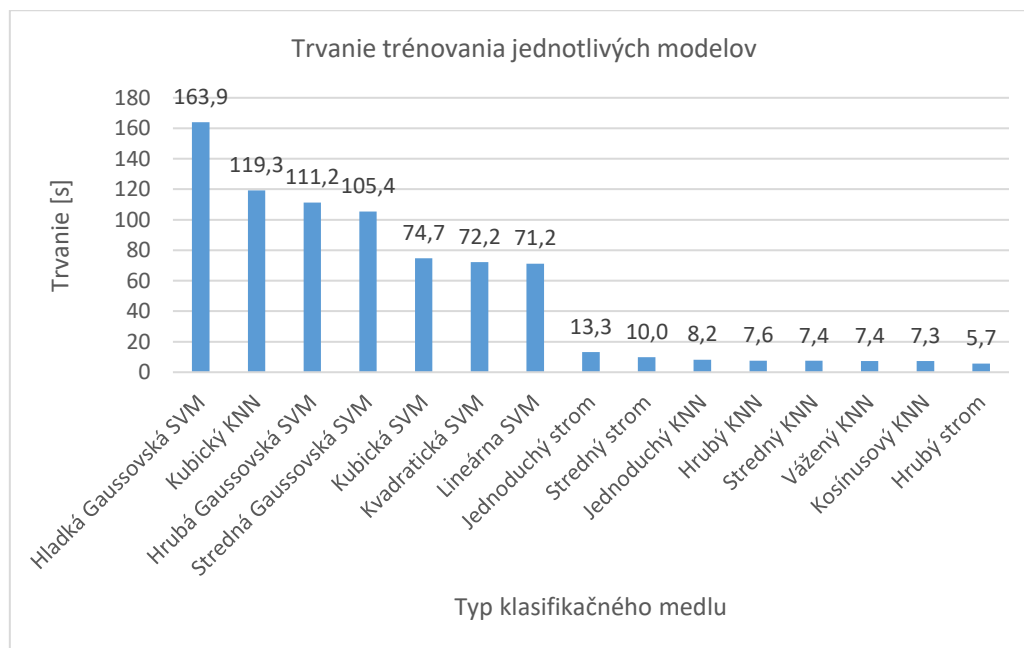
Na klasifikáciu získaných údajov sme použili aplikáciu Matlab Classification Learner (MCL). Táto aplikácia slúži na tréningovanie modelov strojového učenia a pomocou nich klasifikuje údaje. Takto môžeme preskúmať **strojové učenie pod dohľadom** a vstupné dáta. Môžeme takto vyberať prvky, špecifikovať validačné schémy, trénovať a vyhodnocovať presnosť modelov. Vykonali sme experimenty za použitia klasifikačných modelov rozhodovací strom, stroje podporných vektorov (SVM) a k-najbližších susedov (KNN). Pred tréningovaním sme rozdelili údaje na tréningovaciu a testovaciu sadu údajov. Tréningová sada údajov predstavuje 80% z pôvodných 5000 záznamoch o ťahoch písania číslic. Zvyšných 20%, čo predstavuje 1000 ťahov písania číslic sme zaradili do testovacej sady údajov. Obe sady údajov obsahovali údaje od všetkých účastníkov tréningovania [3].

Parametre v oknách sme nastavili ako prediktory a ako očakávanú odpoveď (response) sme nastavili hodnotu AID reprezentujúcu anonymnú identitu účastníka. Validáciu sme nastavili krížovú validáciu s ôsmimi partíciami. Táto validačná schéma je odporúčanou validačnou schémou pre menšie dátové zdroje. Jej princípom je rozdelenie dát na k partícií, pričom model testuje jednou (validačnou) partíciou a trénuje ostatnými. Tento postup sa potom niekoľko krát zopakuje a zakaždým sa vyberie iná partícia ako testovacia z zvyšných slúžia na tréningovanie modelov [8]. Výsledky presnosti krížovej validácie sme zaznamenali spolu s trvaním. *Presnosť* bola definovaná ako *podiel správnych predikcií* ku *celkovému počtu predikcií* krát 100%. Pre každý natrénovaný model sme exportovali funkciu. Vstupom tejto funkcie je súbor s testovacími dátami a výstupom je presnosť testovania. Výsledky presnosti tréningovania a testovania modelov môžeme vidieť na Obr. 4. Trvanie tréningovania jednotlivých typov modelov je zobrazené na Obr. 5 [3].

Vo všeobecnosti platí, že modely typu SVM sú v kontexte našich vstupných údajov najpresnejšie, aj keď hladká Gaussovská SVM dosiahla najnižšiu presnosť, ktorá pri testovacej vzorke dosiahla hodnotu iba 27,7% ale 4 najpresnejšie typy modelov sú modely typu SVM, pričom pri testovaní každý z nich dosiahol presnosť nad 77%, najviac však 79,6%. Pri tréningu dosiahol najvyššiu presnosť model typu kvadratická SVM s presnosťou 91,6%. Nevýhodou SVM je však vysoká výpočtová náročnosť. V našich podmienkach trvalo tréningovanie jednotlivých SVM modelov viac ako 70 s pre jeden model. Tréningovanie najpresnejšej SVM trvalo až 119,3 s a paradoxne tréningovanie najmenej presnej, hladkej Gaussovskej SVM trvalo až 163,9 sekúnd. Najrýchlejšie tréningovanie bolo pri modeloch zo skupiny rozhodovacích stromov, kde tréningovanie navyše trvalo modelu typu jednoduchý strom, pričom trvalo to 13,3 s, pričom presnosť tréningovania tohto modelu dosiahla 64% [3].



Obr. 4 Presnosť tréningovania a testovania jednotlivých modelov [3].



Obr. 5 Trvanie tréningovania jednotlivých modelov [3].

VII. ZÁVER

Medzi behaviorálne biometrické modalities patrí okrem iného aj biometria založená na rukopise. Rukopis zároveň patrí medzi najsilnejšie behaviorálne charakteristiky. Vykazuje dobré vlastnosti, ktoré sú dôležité pre biometrickú modalitu. Je univerzálny, unikátny, permanentný a veľmi jednoducho snímateľný [3].

Prvým krokom v biometrickom systéme je akvizícia dát. Jedná sa o proces získavania dát pomocou senzoru alebo iného akvizičného zariadenia. My sme na akvizíciu dát použili priekopnícke zariadenie Crayonic KeyVault [5]. Akvizície vzoriek sa zúčastnilo 27 subjektov.

Nazbierané dáta sme ďalej normalizovali a parametrizovali. Pracovali sme s tromi verziami údajov. Základná verzia pozostávala časovo ovzorkovaných údajov, ktoré slúžili na analýzu dát v čase, vzorku po vzorke. Časovo ovzorkované dáta sme transformovali do okien a tieto dáta sme následne analyzovali analytickým nástrojom Power BI. Dáta transformované do vzoriek sme nakoniec rozdelili na dva súbory, v pomere 4:1 na trénovací a testovací súbor [3].

Strojové učenie je jeden z možných spôsobov, ako klasifikovať jednotlivca a vyhodnotiť jeho identitu. Na trénovanie modelov sme využili aplikáciu Matlab Classification Learner, ktorá je súčasťou rozšírenia aplikácie Matlab o súbor nástrojov na štatistiku a strojové učenie. Spustili sme reláciu a ako vstupné dáta sme použili trénovací súbor. Parametre v oknách sme nastavili ako prediktory a AID subjektu sme nastavili ako očakávaný výstup. Na odhad výkonu modelov sme použili krížovú validáciu. Na trénovanie modelov sme využili modely typu rozhodovací strom, SVM a KNN. Natrénované modely sme následne exportovali a otestovali testovacími dátami [3].

Najpresnejší zo všetkých trénovaných modelov v našich podmienkach, je model typu kvadratická SVM. Pri trénovaní dosiahol presnosť 91,6% a pri použití nových, pre model neznámych testovacích dát, sme vďaka tomuto modelu dosiahli testovaním presnosť 79,6%. Ako ďalšie dostatočne presné modely sa javia aj modely typu kubická SVM s presnosťou trénovania 91,3%, stredná Gaussovská SVM s presnosťou 89,5% a lineárna SVM s presnosťou 89,0%. Takto natrénované modely, s takouto presnosťou by mali fungovať ako modul porovnávania v biometrickom systéme bezpečnosti využívajúcom rukopis ako biometrickú modalitu [3].

POĎAKOVANIE

Táto práca bola podporená Agentúrou na podporu výskumu a vývoja vďaka projektu APVV SK-TW-21-0002, Kultúrnou a edukačnou grantovou agentúrou Ministerstva školstva vedy výskumu a športu Slovenskej republiky vďaka projektu číslo KEGA 009TUKE-4/2019 s názvom „Inovácia obsahu a príprava učebných textov pre predmet Biometrické systémy bezpečnosti“, Vedeckou grantovou agentúrou realizáciou výskumného projektu VEGA 1/0753/20 „Robustné rečové technológie metódami hlbokého učenia“ a VEGA 2/0165/21 „Technológie automatického spracovania reči na pomoc v krízových situáciách“ financovaných z prostriedkov Ministerstva školstva, vedy, výskumu a športu SR, Asociáciou COST v projekte CA18231 a nakoniec aj cez Operačný program Výskum a inovácie, financovaného z Európskeho fondu regionálneho rozvoja v rámci projektov ITMS: 313011D232 a ITMS: 313011T571.

LITERATÚRA

- [1] Bučko J. Výukový nástroj na multimodálnu biometriu, Bakalárska práca, školiteľ: M. Pleva. Technická univerzita v Košiciach 2020;
- [2] Shaveta Dargan, Munish Kumar, A comprehensive survey on the biometric recognition systems based on physiological and behavioral modalities, Expert Systems with Applications, Volume 143, 2020, 113114, ISSN 0957-4174, <https://doi.org/10.1016/j.eswa.2019.113114> dostupné online 13.5.2022
- [3] Bučko J. Analýza behaviorálnych biometrických dát, Diplomová práca, školiteľ: M. Pleva. Technická univerzita v Košiciach 2022;
- [4] Brunelli R, Falavigna D. Person identification using multiple cues. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1995;17(10):955-966 <https://doi.org/10.1109/34.464560> dostupné online 13.5.2022
- [5] <https://www.crayonic.com/> dostupné online 20.1.2022
- [6] <https://www.techbyte.sk/2021/03/zabudnite-hesla-slovaci-crayonic/> dostupné online 20.1.2022
- [7] Gomez-Barrero M, Galbally J, Morales A, Fierrez J. Privacy-preserving comparison of variable-length data with application to biometric template protection. IEEE Access. 2017;5:8606–19.
- [8] https://www.mathworks.com/help/stats/select-data-and-validation-for-classification-problem.html?s_tid=srchtitle_Select%20Data%20and%20Validation%20for%20Classification%20Problem_1 dostupné online 6.4.2022

Training and testing system for a humanoid robot in a virtual reality

¹*Branislav SOBOTA*, ²*Miriama MATTOVÁ*, ³*Róbert FEDORCO*, ⁴*Štefan KOREČKO*

^{1,2,3,4}Department of Computers and Informatics, Faculty of Electrical Engineering and Informatics, Technical university of Košice, Letná 9, 04200 Košice, Slovak republic

¹branislav.sobota@tuke.sk, ²miriama.mattova@tuke.sk

Abstract — Main part of paper is focused to description of the training and testing system for the movement of a humanoid model of a robot simulation. This process takes place using a program with virtual reality support (using LIRKIS GCVE platform) and also by creating our own robot model. The result of the work consists in the creation of simulated motion data by combining the functionality of the model and implemented components in our virtual 3D space. We verified the correctness of the output files defined by the URDF structure, which is used in most humanoid models. The process itself and testing took place in the virtual scene we created.

Keywords — humanoid robot, virtual reality, A-frame, ROS reality

I. INTRODUCTION

Humans are primarily social beings that spend a lot of time being involved in interactions with other humans. Understanding the underlying cognitive and neural mechanisms during these social interactions (second-person approach) is important, since social cognition is fundamentally different when we are in interaction with others rather than merely observing them [23]. Two neuroanatomically distinct large-scale networks have gained center stage as the neural substrates of social cognition: mirror neuron system (MNS) and the mentalizing network. The former is believed to give us a “first-person grasp” of the motor goals and intentions of the observed individual [21]. The latter is believed to yield an inferential, reflective, and what might be called a third-person grasp of others’ mental states [10].

Human–human interactions are of different types and in modern society they are gradually being extended to interactions between humans and robots (human–robot interaction, HRI), covering three “co-actions” (cooperation, communication and collaboration). In particular, robots with humanoid shape have been proposed as a plausible controllable tool for studying social interactions [24]. The studies focusing on human MNS are numerous and the related mechanisms can be replicated in cognitive robots. At the same time, humanoid robots have been demonstrated to play an increasingly important role in human society, with their applicability across many tasks or contexts [4].

HRI has demonstrated a rapid progress in the last decade [12] as a promising new field offering numerous applications ranging from industrial collaborative scenarios, through educational school settings for children, to domestic environments for elderly people, in which a socially-intelligent robot plays an important role [6]. A core subfield within HRI is represented by human-robot collaboration [5] where both partners aim to achieve a common goal in real time. This requires reaching a deeper level of interaction, shared understanding, and coordination [8].

Given the physical properties of robots, one key feature of HRI is trust [16], which is considered a social glue that connects people and promotes collective goals. Where humans and robots are working as partners, the performance of work can be diminished if humans do not trust robots appropriately. Achieving trust in HRI rests on a number of factors [20] and entails certain requirements when it comes to the robotic movement: on the one hand, its effector must move safely [2] and as expected, that is, in a predictable way. On the other hand, the robot is also faced with another, often more critical burden of conveying its intent [19]. This can be achieved by legible motion, that is “readable” [25], “anticipatory” [11], or “understandable” [3]. It appears that predictability (goal-to-action) and legibility (action-to-goal) are fundamentally different and often contradictory properties of motion [9]. Hence, they can be optimized (in a learning robot) or tested individually (e.g. by human observers). [9] presented pilot experiments along this direction and proposed formalization of the approach allowing control of the robotic design and

applying quantitative measures for dependent variables. Transparent HRI supports the capability to read the partner's intentions that leads to the theory of mind, involving the mentalizing network [27]. An equally important aspect is the safety of robot actions for the human partner [18].

Virtual reality (VR) represents an advanced technology allowing the designers to create rich virtual environments with 3D visual scenes and objects. VR can be defined as a type of user-computer interface that implements real-time simulation of an activity or environment allowing user interaction via multiple sensory modalities [1]. VR simulators allow exposure to realistic and challenging environments through the presentation of dynamic graphics and audio, while also maintaining safety and control. VR offers a wide range of applications, for example, in entertainment, art, design, neurorehabilitation, and others. Typically, VR is enabled by a head-mounted display (HMD) which is also the method used.

Last but not least, using VR offers a number of advantages over physical HRI: high flexibility in environment setup, good control of independent variables, safety, and possibility for teleoperation, to name a few [28]. VR is becoming realistic and has become more usable with regards to research purposes, including the replication of physical HRI studies. However, this only works when the participant is sufficiently immersed in the VR environment. There exist solutions with mostly good immersion, but a number of challenges remain [7].

VR and related technologies have been developed in the authors' home laboratory (LIRKIS laboratory at DCI FEEI TU Košice) for several years. Also for the context of this paper was developed LIRKIS Global - Collaborative Virtual Environments (LIRKIS G-CVE) [14]. LIRKIS G-CVE is a web-based system, built on top of the A-Frame¹ web VR framework and its extension, called Networked-Aframe² (NAF). Thanks to A-Frame, it is possible to run LIRKIS G-CVE on various user devices, from desktop computers and notebooks through tablets and smartphones to virtual and mixed reality headsets. The integration of NAF makes the VE global and collaborative by allowing on-line multi-user interaction in real-time [13]. LIRKIS G-CVE adds several reusable components, particularly the Smart-Client Interface (SCI) and the Enhanced Client Access (ECA) layer. SCI allows the use of tablets and smartphones as input devices. Thanks to ECA, a client may access detailed information about the avatars, representing other clients/robots in the same shared virtual environment. In the following will be short described virtual robot model and the pilot application using LIRKIS-GCVE for visualisation a basic movement simulation of humanoid robot (this model).

II. ANALYSIS OF AVAILABLE SOLUTIONS

The pilot application is created using the LIRKIS-GCVE platform with the web framework called Glitch. The main function of the application is to simulate motion, namely of some parts of a humanoid robot model. This simulation/visualization has the task to calculate the rotation and motion vector of the effectors and apply them to a fixed object that of the robot NICO model [15] (

Fig. 1). NICO was chosen mainly because of its open development platform. This is done in real time and with a precise data structure. The application consists of scene and entity models. These models represent the user and the robot model on which the simulation is running. The application has keyboard input and mouse control providing movement. The task of the application user is to record and therefore simulate this movement in a shared virtual environment.

Currently, several platforms and languages are used for the development of humanoid models, from design to control and implementation of add-ons. Whereas these are more complex data models to work with, the language where these models are developed and controlled is mostly Python. It offers a wide range of additional libraries that can be easily implemented to evolve the model. Since this concept models is designed for incremental machine learning, this language is an ideal candidate for this work. As far as model design is concerned, software solutions such as SketchUP and Autodesk are used. However, these do not offer such a wide range of possibilities for manipulation as Blender. On the other hand, the latter offers the layout of individual elements from the skeleton to the texture of the models, where each one of the elements has multiple rotation, position, weights and particle modifiers defined area and vertices of the object. Regarding the development and manipulation platform of a given of the model, the solution is the LIRKIS GCVE platform. This platform offers a simple solution for the collaborators' work using the Glitch web interface. It is platform based on 3D object and entity modelling using HTML and Javascript.

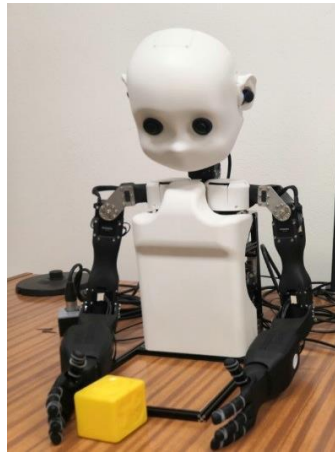


Fig. 1 NICO robot

A. Used output format of software applications

As a result, there are already several open-source systems for simulating humanoid and commercially produced robots. Some are outdated and use only limited hardware tools such as mouse and keyboard. Others, on the other hand are adapting to the new hardware available for VR support system integrations (HMD, HTC, Oculus...) such as CoppeliaSim [22]. This project represents probably the biggest opportunities in terms of development as an open-source platform. Other solutions represent a paid option for commercial use and access to the source code and architecture of the software system is either strictly limited or completely secured. Also, most simulators are software that need to be run locally and thus downloaded. From other point of view, this solution provides more accurate simulation results and faster computational response. Interestingly, most of the mentioned simulation solutions are built on a system ROS (Fig. 2) [26]. In accordance with this system, the extensive use of data files such as JSON, PNG, and the established BSON format file format for faster data transfer. These formats have the least load on the network and hence faster response time between the model and user commands. ROS Reality represents a bridge between the VR compatible headset and control program e.g. the Unity. It can also identify robot manipulation tasks.

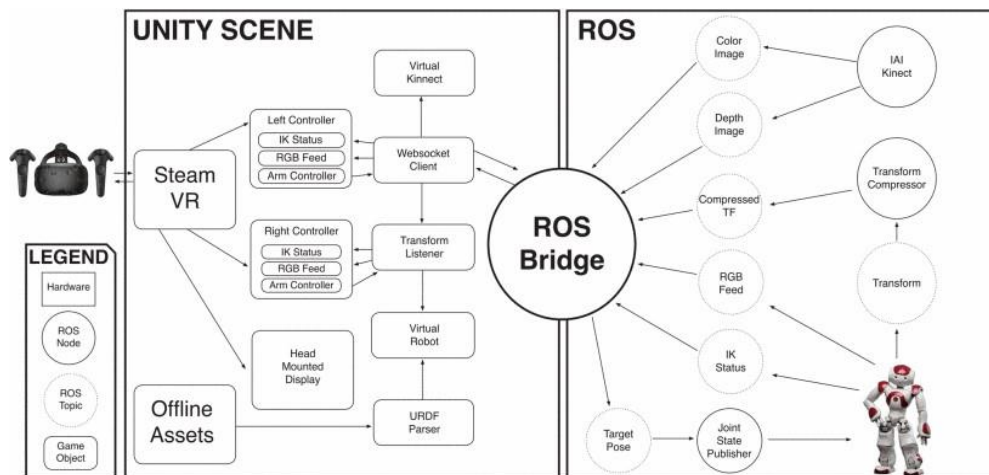


Fig. 2 ROS system

III. DESIGN OF MODEL

The model itself was designed to simulate humanoid robots. Therefore, the basic parts consist of a body, an arms and a head. Different models were available on the internet, but we decided to create our own, since the free models available did not have the correct structure of the joints and their internal distribution within the model. For it we used the Blender tool, and used the parent-child joint structure as an example URDF joints. These describe that each joint must be directly dependent on the previous while parent links can be both children and children of themselves parents of other links in the model. Our model contains a base parent, three "sub-parents" (they are both parents and inherit traits) and two offspring. This structure is used in the modelling tool is referred to as "Armature" (Fig. 3) and the individual parts are labeled as the skeleton. Each this

skeleton has at least one motion joint defined. Since the A-frame platform only supports model formats like obj, gltf or glb so in our case we modified the model by adding textures and exported this model to gltf format. Directly on the LIRKIS GCVE platform, the model is added as a whole and by adding the gltf model component. For better functionality and smooth simulation, the model contains only meshes of type

- *cone*
- *cube*
- *cylinder* with a specific skew modifier added to each type to determine the exact number of vertices and edges. The model texture was selected from textures:com=library and used in the application as the principle two-way scattering function. This type of texture deposition is directly supported by the A-frame platform and can be adjusted with parameters such as roughness, hue or specularity. When retrofitting the surface and mainly by adding a mirror-like material, a lower frame rate per second in the scene is achieved.

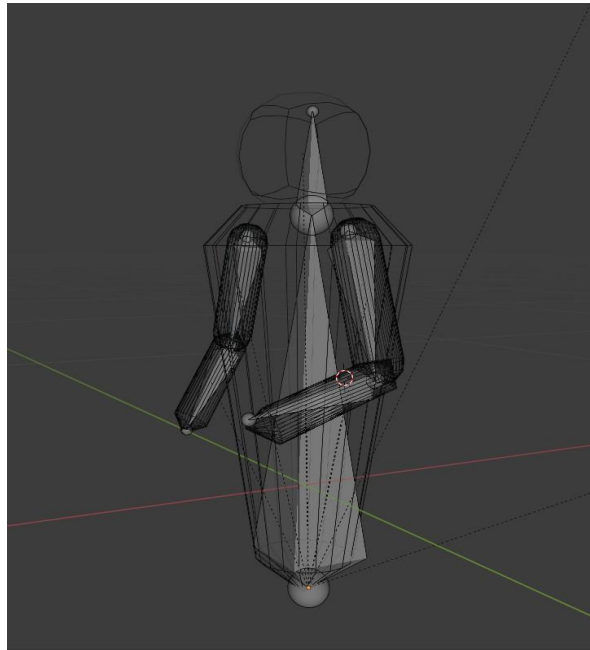


Fig. 3 Robot model with „Armature“ kinematic structure

A. Recording process of model simulation

There is recording of two objects in a scene. The first object on which simulation and testing is performed is the humanoid robot model itself. The second is the avatar itself. The process is interactive and start-off recording of one or the other object is independent of each other. The functionality of one of the process does not depend on the other and thus the following possibilities arise

- simultaneous recording of the model and the user
- recording of the model only
- recording of the user only
- recording process switched off

We decided to record the simulated motion by running the created process in the scene. This functionality is triggered by the user's input interaction, namely by pressing the "M" key. The recording is visualized in the scene by a red flashing indicator on the right side of the screen. The recording interacts directly with the motion simulation functions, i.e. all data such as rotation, selected part of the model and timestamp are stored in their own fields only when the recording process is enabled. To turn the recording off, press the same key input as when it was started. The result is a file of data type JSON whose size depends on the number of moves. Given that each movement and rotation is recorded, the size of the resulting file will be affected by they number. The more inputs from the user the larger the file size. The structure of the output data is as follows

- *the name of the part of the model* being captured
- *time stamp*
- *x,y,z object rotation coordinates*

The robot model is inactive within the motion in the scene therefore the coordinates of the position are not written to this structure.

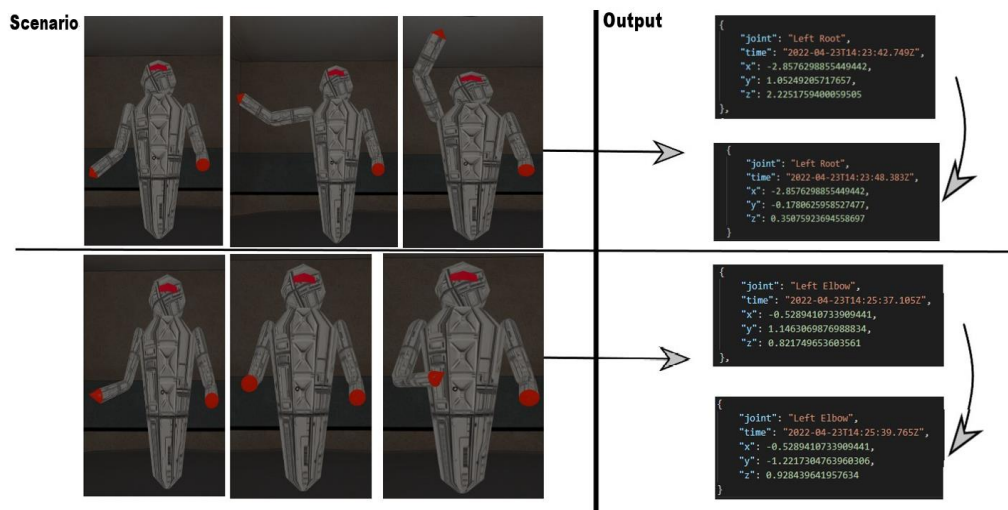


Fig. 4 Example of recording of interaction of robot „body“ and output

B. Recording process of player activity

Similar to the robot object, this approach is based on input interaction by pressing the "P" button. During recording, every movement and rotation is sensed of the user's (his avatar) camera. Since the camera is a direct component of the avatar and is located in the head area, the output parameter is the position and rotation of the user's / observer's head. However, in this case the following parameters are stored

- *position*
- *rotation*
- *time stamp*

The difference from recording the robot model is the removal of the rotation at the x, y, z coordinates and adding the position within the scene. An avatar is a moving object within the scene at a horizontal 2D surface so its X coordinate (vertical up-down movement) does not change. The output file is automatically saved to the local disk of the device on which the recording takes place under the name

- KPI_Record_date, time of upload - if it is a JSON robot model file or
- user-recording date, time of recording - if it is a user's JSON file.

In both cases, the same principle is used when recording both the player and the model, namely a format that uses human-readable text to store and transfer data objects consisting of attribute/value pairs and fields (or other serialize values). It is a common data format with a variety of uses in electronic data interchange, including the interchange of web applications with servers. In our case, we used the binary large object ("BLOB") structure.

IV. EVALUATION

To evaluate the functionality of the program and the training system, we have prepared a concrete scene enriched with a building. The inner part of the building contains the base on which a humanoid model is placed. The room was selected from the <https://hubs.mozilla.com/spoke/projects/create?filter=remixable> where various *glb* and *gltf* format scenes and objects are available. The model is fixed in the scene and cannot be moved. For better visibility of the end joints, we have marked the end of their textures in red (see also Fig. 4). The environment and the background are matched to the colors of the building. At the left bottom of the screen (Fig. 5), a scale model of the robot is shown, with a description of the keyboard shortcuts for selecting the different parts of the joints. If the user presses one of the keys, the user can verify the functionality of the joint by moving the cursor. Above this view is shown to start or stop the recording of either the user or the humanoid robot. The main scene is divided into three parts (Fig. 5). The first part contains the user's basic camera. This camera is the only moving object of the program and the ability to freely observe the humanoid model. The other two parts are consist of secondary cameras. These are set to a frame rate of 30 frames per second. This can be set in the *spectator.js* file by changing one number. In evaluating the output of this work, the following steps were followed

- recording output
- joints movement
- functionality in the scene

The most important point of this work was the correctness of the output of the recording processes. These outputs were in the form of two files. The first data file was a recording of the humanoid model and the second file was a recording of the user's / avatar's camera. For our needs, the URDF joint structure was used. The second point was the correctness of the movement of individual joints. This functionality was implemented by the interaction of the cursor and keyboard shortcuts in the virtual scene. The process consisted of the movement of the selected joint itself, but also the simultaneous influence of the parent joint or its child joint.

The last/final point was the process of the mentioned application functionalities implementation. The process was tested in a virtual scene in the role of an observer. All tests were performed sequentially. First, the movement functionality of the model was tested. First of one joint and the possibility of selection was subsequently added. Based on this implementation, we designed the recording process and incorporated it into the motion process. Finally, we checked the functionality of these processes and the result was a file containing an accurate joint description and their individual rotation in 3D space.

5 people (age 20-24) took part in the testing. All participants performed testing in the following web browsers: Google Chrome (fps in the range 10-40), Mozilla Firefox (fps in the range 10-40) and Opera GX (stable fps 60 for all test participants).



Fig. 5 Example of virtual environment with humanoid robot model

V. CONCLUSION

Virtual reality allows to create rich virtual environments with 3D visual scenes and objects to test their capabilities and parameters even before their physical creation. VR can thus represent a user-computer interface that implements a real-time simulation of an activity or environment. The use of the presented pilot application provides the possibility of testing and recording the movements of the humanoid robot model and subsequently implement it in a physical humanoid robot. This reduces the cost of developing the respective type of humanoid robot. In future work, we plan to devote ourselves to the implementation of a broader formalism for the description of the robot and possible interfaces for the application of artificial intelligence.

ACKNOWLEDGMENT

This work has been supported by the APVV grant no. APVV-21-0105 "Trustworthy human-robot and therapist-patient interaction in virtual reality".

REFERENCES

- [1] Adamovich S., Fluet G., Tunik E, Merians A. (2009). Sensorimotor training in virtual reality: a review. *NeuroRehabilitation* 25: 29-44.
- [2] Alami R. et al. (2006a). Safe and dependable physical human-robot interaction in anthropic domains: State of the art and challenges, IROS Workshop on pHRI.
- [3] Alami R. et al. (2006b). Toward human-aware robot task planning. In *AAAI Spring Symposium*, pp. 39-46.
- [4] Bahishti A.A. (2017). Humanoid Robots and Human Society. *Advanced Journal of Social Science*, 1(1), pp. 60-63
- [5] Bauer A., Wollherr D., Buss M. (2008). Human-Robot Collaboration: A Survey. *International Journal of Humanoid Robotics*, 5(1), pp. 47-66.
- [6] Dautenhahn K. (2007). Socially intelligent robots: dimensions of human-robot interaction. *Phil. Trans. of the Royal Society of London*, 362(1480): 679-704.
- [7] Dianafar M., Latokartano J., Lanz M. (2020). Review on existing VR/AR solutions in human-robot collaboration. *Procedia CIRP* 97, 407-411.

- [8] Dillenbourg, P. (1999). What do you mean by collaborative learning? In *Collaborative-learning: Cognitive and Computational Approaches*. Elsevier.
- [9] Dragan A.D., Lee K., Srinivasa S.S. (2013). Legibility and predictability of robot motion. *ACM/IEEE International Conference on Human-Robot Interaction*, pp. 301-308.
- [10] Frith U. & Frith C.D. (2010). The social brain: Allowing humans to boldly go where no other species has been. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 365(1537), 165–76.
- [11] Gielniak M., Thomaz A. (2011). Generating anticipation in robot motion, in *RO-MAN*, pp. 449-454.
- [12] Goodrich M.A., Schultz A.C. (2008). Human-Robot Interaction: A Survey. *Foundations and Trends in Human-Computer Interaction*, 1(3), pp. 203-275.
- [13] Hudák M., Korečko, Š., Sobota B. (2019). Enhancing team interaction and cross-platform access in web-based collaborative virtual environments. In *IEEE 15th International Scientific Conference on Informatics*, pp. 171–176.
- [14] Hudák M., Korečko Š., Sobota, B. (2020). Advanced user interaction for web-based collaborative virtual reality. In *11th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*, 343–348, doi:10.1109/CogInfoCom50765.2020.9237899.
- [15] Kerzel M., Pekarek-Rosin T., Strahl E., Heinrich S., and Wermter S., “Teaching nico how to grasp: An empirical study on crossmodal social interaction as a key factor for robots learning from humans,” *Frontiers in Neurorobotics*, vol. 14, 2020. [Online]. Available: <https://www.frontiersin.org/article/10.3389/fnbot.2020.00028>
- [16] Khavas Z.R. (2021). A Review on trust in human-robot interaction. arXiv:2105.10045v1
- [17] Korečko, Š., Hudák M., Sobota S., Marko M., Cimrová B., Farkaš I., Rosipal R. (2018). Assessment and training of visuospatial cognitive functions in virtual reality: proposal and perspective. In *9th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*, pp. 39–44, doi:10.1109/CogInfoCom.2018.8639958.
- [18] Lasota P.A., Song T., Shah J.A. (2017). A Survey of Methods for Safe Human-Robot Interaction. *Now Foundations and Trends*, <https://doi.org/10.1561/23000000052>
- [19] Lichtenthäler C., Lorenz T., Kirsch A. (2011). Towards a legibility metric: How to measure the perceived value of a robot, in *ICSR Work-In- Progress-Track*.
- [20] Nam C., Lyons J. (2020). *Trust in Human-Robot Interaction* (1st ed.), Academic Press.
- [21] Rizzolatti G., Sinigaglia C. (2010). The functional role of the parieto-frontal mirror circuit: interpretations and misinterpretations. *Nature Reviews Neuroscience*, 11(4), 264-270.
- [22] Rooban S., Suraj S. D., Vali S. B., and Dhanush N., “Coppeliasim: Adaptable modular robot and its different locomotions simulation framework,” *Materials Today: Proceedings*, 2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S221478532100078X>
- [23] Schilbach, L. et al. (2013). Toward a second-person neuroscience. *Behav. Brain* 36, 393–414.
- [24] Sciutti A., Sandini G. (2017). Interacting with robots to investigate the bases of social interaction. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 25(12), 2295-2304.
- [25] Takayama L., Dooley D., Ju W. (2011). Expressing thought: improving robot readability with animation principles. In *Human-Robot Interaction*.
- [26] Whitney D., Rosen E., Ullman D., Phillips E., and Tellex S., “Ros reality: A virtual reality framework using consumergrade hardware for ros-enabled robots,” in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2018, pp. 1–9.
- [27] Vinanzi S., Cangelosi A., Goerick C. (2021). The collaborative mind: intention reading and trust in human-robot interaction. *Iscience*, 24, <https://doi.org/10.1016/j.isci.2021.102130>
- [28] Wijnen L., Lemaignan S., Bremner P. (2020). *Towards using virtual reality for replicating HRI studies*. HRI'2020 Companion, Cambridge. UK.

Data flow analysis of videoconference devices in network infrastructure

¹Martin MURCKO, ²František JAKAB, ²Dávid CYMBALÁK,
²Roman VÁPENÍK

¹Department of Computers and Informatics, Faculty of Electrical Engineering and Informatics Technical University of Košice, Slovak Republic

²Central management office of National telepresence infrastructure, Slovak Centre of Scientific and Technical Information, Slovak republic

martin.murcko@student.tuke.sk, frantisek.jakab@cvtisr.sk, david.cymbalak@cvtisr.sk,
roman.vapenik@cvtisr.sk

Abstract – This paper explains the importance of monitoring network communication, describes various types of anomalies and several methods to detect them. It deals with the topic of using exponential smoothing algorithms for creating forecast and the possibility of their usage for dynamic monitoring of networks. Based on this analysis, a design and subsequent implementation of an experimental real-time monitoring system was created. Its functionality was demonstrated in the analysis of network traffic in a model network with simulated videoconferencing.

Keywords – ClickHouse, data analysis, exponential smoothing algorithms, Kafka

I. INTRODUCTION

The growth in the number of users of the internet was doubled every year [1] compared to the previous year in the beginnings of 21st century. After this massive expansion the growth of the Internet of Things (IoT) started. Smart devices have become more affordable and accessing the Internet is easier than before thanks to mobile networks with much higher speed rates than in the past. Amounts of transferred data are rising [2] due to the usage of cloud environments, social networks and streaming services. Based on these factors, it is more difficult to ensure the quality of services, it is harder to make the right decisions [3] when designing networks.

The need for precise network monitoring is greater than ever before. It helps with fast troubleshooting, network security, learning about different protocols, preventing accidents or alerting when some anomalous event occurs in the network.

II. NETWORK MONITORING

Network monitoring consists of three consecutive steps [4] - collecting data, their transformation to needed format and finally analyzing them. There are multiple methods [4] how to collect data. Hubbing Out is the easiest one. All you need to is to separate monitored part of the network with a hub and after that monitor the traffic on this hub only. Using a tap is placing a special piece of hardware between two points in the network, and it will catch all packets flowing through it. Using an ARP cache for collecting network data is used mostly by hackers and it consists of pretending to be a device with specific MAC address, for which the original sent message was intended. The most used method of capturing traffic is port mirroring. It consists of copying data from one port to another and no additional piece of hardware is needed for this case.

Captured traffic has the form of Protocol data units cached on the third layer of the OSI model as IP packets. They contain information like destination and source IP address, number of TCP or UDP ports etc..

III. ANOMALY DETECTION AND DEFINITION OF ANOMALIES

Anomaly detection consists in searching for patterns [5] in the data that do not match expected normal behavior, which often signals that there is something wrong in the system. Anomalies could be caused by performance issues, datacenter failures or security breaches in the network.

Network anomalies could be divided into these three categories [6]:

- 1) Network Operation Anomalies - broken infrastructure, power outages etc.
- 2) Flash Crowd Anomalies - high amount of traffic due to a big event on some website like a sale etc.
- 3) Network Abuse Anomalies - for example DDoS attacks.

The detection itself consists of defining [7] what is normal behavior and what is anomalous. It is considered as a hard task, because there could be many edge cases that can skew results when defining what is normal and what not and anomalous activities like security attacks in the network have often the ability to adapt their behavior and stay hidden. Most of the existing techniques used in the field of anomaly detection can work only on a particular problem set, there is no general approach. That is the cause why many fields of studies like machine learning or data mining are used for this purpose.

Anomalies can be divided into three categories according to their characteristics [7]:

- Point Anomalies - in contrast with the rest of data is an individual data instance considered as anomalous.
- Contextual Anomalies - an individual data instance is not anomalous at all, but when the context is taken into account, it is considered as an anomaly.
- Collective Anomalies - an individual data instance is not anomalous at all, but only when an entire dataset or collection of related data instances is anomalous.

An example of the graphical representation of individual types of anomalies can be seen in the Fig. 1

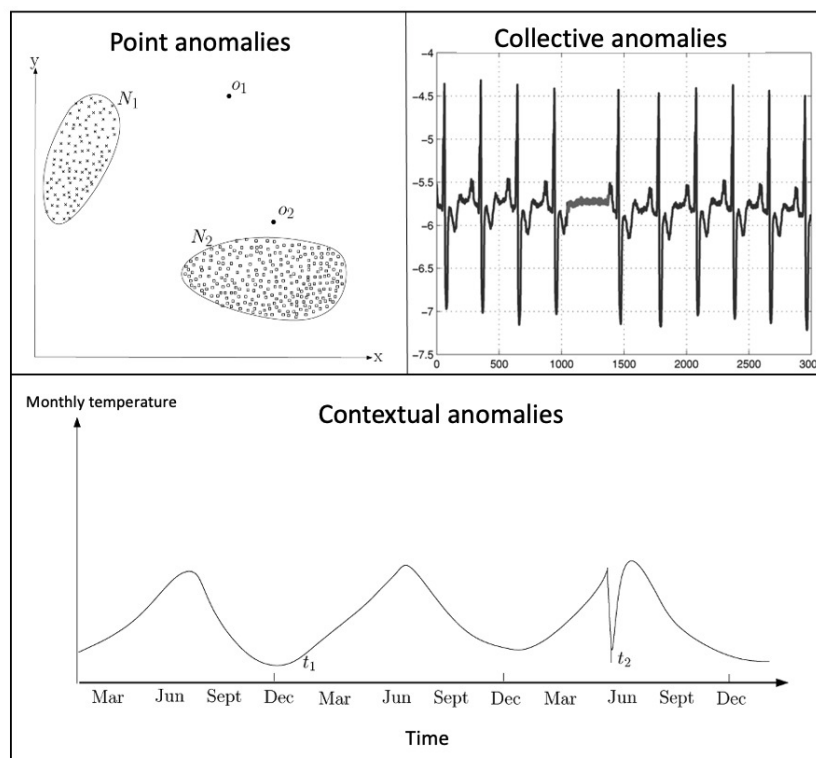


Fig. 1 Graphical interpretation of anomalies by their characteristics [7]

Techniques used for for training a model that will detect anomalies [7]:

- Supervised - every data instance has a label whether the data is anomalous or not.
- Semi-supervised - only data instances labeled as normal are used.
- Unsupervised - data instances have not any labels. This technique is working with an assumption, that normal instances are the most common in the provided dataset.

These techniques are used by various methods. Some of them have been established as a standard in this field of study:

- Classification methods - decision trees, support vector machines etc.
- K-nearest neighbours method.
- Clustering method .
- Methods using statistic models - moving average, statistical z-score, regression - ARIMA (Auto Regression Integrated Moving Average).
- Histogram method.

It is easier to monitor networks in a stable environment, where Quality of Service is high and packet delivery rate is more than 99%. In networks with lower reliability, it is more difficult, because the definition what is normal and what is an anomaly can change very often. In these cases, the best solution is to use statistical prediction models. Forecasts are made using time series data [8], which can be characterized as a set of observations, each at a specific time so every observation in this set has a unique timestamp. Time series data could be decomposed into four components [9]:

- 1) Level - the average value for a specific time period,
- 2) Growth - the average increase or decrease of the value over a period of time,
- 3) Seasonality - a pattern that repeats itself with a fixed periodicity,
- 4) Error - unexplainable white noise.

When a prediction is made at time t it is simply compared with the observed value at that time and a decision is made if it is an anomalous data instance or not. Other use case consists only of trying to find anomalous instances in forecasted data and predict an event that could happen in monitored network.

The most straightforward [10] statistical method for creating forecast is the naive method. It assumes that the predicted k -th element is equal to actual observed value \pm standard deviation. Neuron networks can also be used in making predictions. A multilayer perceptron is often used for this task. Advantage of neuron networks is that there is no need to adjust complex parameters when creating a forecasting model, but training them could be very complex. Anomaly detection method ARIMA could be also used for forecasting. If time series data has signs of growth or seasonality, complex adjustments must be made to make this autoregressive method work in the right way.

Exponential smoothing [11] is one of the oldest methods used for forecasting created in the 1950s. It is easy to use, not so complex and yields good results. Historical values from given time series are used with combination of weighted arithmetic mean. The greatest weights are on the most recent data instances and the weights are decreasing exponentially to the older ones. Simple exponential smoothing is used for time series without or minimal growth and seasonality. It takes into accounts only the level component and according to set α parameter called smoothing parameter it assigns weights to individual data instances and manages how fast they decline towards older values in series. For α the condition $0 \leq \alpha \leq 1$ applies.

The forecast equals the result for smoothing formula l_t for $t > 0$:

Forecast:

$$\hat{y}_{t+h|t} = l_t$$

Level component:

$$l_t = \alpha y_t + (1 - \alpha)l_{t-1}$$

Exponential smoothing that takes into account the growth component of time series is called double exponential smoothing or Holt's linear trend method. In addition to the α parameter it uses the β parameter describing the trend. Similar as for α , for β the condition $0 \leq \beta \leq 1$ applies. The formula for double exponential smoothing forecast has the following form:

Forecast:

$$\hat{y}_{t+h|t} = l_t + hb_t$$

Level component:

$$l_t = \alpha y_t + (1 - \alpha)(l_{t-1} + b_{t-1})$$

Growth component:

$$b_t = \beta(l_t - l_{t-1}) + (1 - \beta)b_{t-1}$$

IV. NETWORK MONITORING TOOLS AND PROGRAMS

When creating a network monitoring system, choosing the right tools and programs is a very important part of the development process.

A. Packet sniffing tools

Packet sniffing tools are using libcap [12] interface to capture and store packets in pcap format or the newer specification pcapng. For further processing with additional tools or applications it is more convenient to use JSON format. There are many tools like tcpdump, Wireshark and its command line version - tshark or addons for programming languages like Scapy for python.

Almost every Linux or macOS distribution comes with command-line packet analyzer called tcpdump [13]. It is developed together with the libcap interface and provides a convenient CLI tool that can be used on remote servers for fast troubleshooting of different kinds of problems.

For more complex monitoring a more appropriate tool is Wireshark [4] or its command line version tshark. It has support for hundreds of protocols and in contrast to tcpdump more complex filters could be created for network monitoring.

Scapy [14] is a packet manipulation program used together with python programming language. Sniffing packets directly in Python opens many possibilities of analysis with the usage of powerful libraries like NumPy or pandas. But the analysis cannot be done in real time and a low throughput rate could be a caveat if bigger parts of network are being monitored.

B. Packet processing

After capturing packets, a tool or database is needed, that will be able to save and load them when performing various analyzes. There are many systems for processing huge amounts of data like packets in the network.

The easiest way is to save captured packets in a pcap file [15] and analyze them using packet sniffing tools. But this method encloses the entire analysis completely into the environment of packet sniffing tools. A more convenient way is to store packet data in SQL or NoSQL databases like PostgreSQL, MySQL, Oracle or mongoDB. The advantages of this approach are high reading speeds and ease of use when using SQL or NoSQL syntax. The disadvantages are the need for a lot of disk space to store hundreds of data instances and analysis cannot be done in real time due to the need of transforming a pcap file to database and only after that the analysis can start.

A more complex approach is to use message-broker Apache Kafka [16] to stream packets. Kafka consist of client and server part. The purpose of the server part is to store produced data in brokers and provides Kafka Connect service for connecting various applications. The client side provides Kafka Producer and Consumer APIs to produce or read streams of events. With Kafka, packets can be streamed and analyzed in real time using any supported application or programming language, for example streaming and storing packets into an analytical database.

C. Analytical databases

Online analytical processing [17] (OLAP), is an approach how to analyze huge amounts of data. This approach requires that most of the requests should be read-only, updates are made in batches and readings of thousands to millions rows should be done in a short time period. It uses column-oriented storing of data which yields up to 100 times faster results than classical row-oriented approach used by relational databases like MySQL, Oracle or PostgreSQL. Analytical databases are using OLAP approach and one of the newest addition in this area is ClickHouse database with much higher speeds [18] than other solutions in this field.

D. Time series databases

Time series databases are specially designed to store time series data. For example, forecasts created by double exponential smoothing when monitoring some network can be stored in such type of databases. Two of the mostly used solutions are Prometheus and InfluxDB. Prometheus [19] uses pull model to retrieve time series metrics by itself in specified intervals. InfluxDB [20] uses push model so the monitored system must push the data itself into the database.

V. DESIGN AND IMPLEMENTATION OF APPLICATION FOR NETWORK MONITORING

After analyzing different tools and applications related to network monitoring and working with huge amounts of data, decisions were made which ones should be used when creating an own real-time monitoring system. For packet capture the best tool is tshark, due to its robustness and complexity so it can cover many use cases. For packet processing the best option is to stream captured packets to Kafka so tshark will have the function of a Kafka producer. Using Kafka does not create any restrictions on what tool or application should be Kafka consumer for produced events in the form of captured packets. For storing packets ClickHouse database turns out to be the best option. It is an analytical database that is much faster than other solutions, and it provides

several tools like consuming Kafka produced events into a special Kafka table or creating views with aggregated records.

To allow users to create custom network monitoring configurations that will be monitored and check created alerts according to them an application will be made. They should be able to specify anomalous states in the configurations for given combination of properties, choose from various monitoring options and check detailed alerts with graphs created by monitored configurations to better understand what happened in monitored network.

In final implementation all supporting applications are running as docker containers. In the beginning packets are captured by tshark, transformed with jq tool to corresponding JSON format and then feed to Kafka. Stream of events is consumed by ClickHouse database in a special table created with KAFKA engine. Other tables are storing packets aggregated for 1,2,5 seconds and for 5 minutes where source table is the Kafka one and data are aggregated and stored via created views.

Frontend application provides for the users the functionality of creating configs that should be monitored and viewing created alerts for them if an anomalous event occurred. For creating configs, following fields can be set:

- Source and Destination IP addresses, TCP Source and Destination ports, UDP Source and Destination ports - packets with these properties set should be monitored.
- Monitoring method - which method should be using when making the analysis. There are four methods:
 - Threshold - monitoring of crossing a threshold, for example 100 packets.
 - Percentage - monitoring of exceeding the fixed percentage rate from total traffic in given period.
 - No traffic - special case of fixed threshold, monitored is the presence of any network communication.
 - Dynamic Prediction - performs forecast based on historical data using the Holt linear method, then the result is compared with the measured value and the percentage difference is computed.
- Aggregation interval - there are two options:
 - Entire interval - selected method is used for entire selected period.
 - Every datapoint in interval - selected method is used for every datapoint in selected period in intervals of 1,2 or 5 seconds.
- Period - monitored time interval.
- Threshold - monitored value, could be a numeric or percentage rate.

Configs are handled by the backend part of created application. It all starts with cron scheduler that loads all configs in specified time intervals and after that analysis method is chosen according to specified monitoring method in config. Forecasting method uses InfluxDB to store prediction datapoints and both alerts and configs are stored in PostgreSQL.

VI. PRACTICAL MONITORING OF VIDEOCONFERENCE ENVIRONMENT

Created solution is able to detect the smallest changes in monitored networks so it is suitable for monitoring networks with videoconference communication. A file with captured videoconference communication, shown in Fig. 2, was used to test the experimental implementation.

In the figure, it is possible to clearly determine the decrease in the total number of incoming packets, approximately from 135th to 165th second, due to turning off the microphone. After it was turned on again, the number of packets stabilized again. In graph of incoming communication, a significant increase in the number of received packets is evident, at about the 200th second, due to sending packets of video communication from the webcam.

For detecting anomaly two configurations were created. Both of them used the Dynamic Prediction method together with aggregation interval Every datapoint in interval option with five second intervals. The first one has as the source IP address set the IP address of the videoconferencing device, the other one has it set as destination IP address.

In Fig. 3, an alert is created, based on the monitored network communication, where the destination IP address is set to 170.72.228.37. The difference in prediction is already visible in the first seconds of communication due to the start of the call, then the communication is stable and at about 9:32:10 another difference in measured between the predicted and monitored value due to anomaly - microphone was turned on. The system also responded to turning off the microphone and a difference of more than 5% was between the predicted and monitored value.

In Fig. 4, an alert is created, based on the monitored network communication, where the source IP address is set to 170.72.228.37. Also as in previously created alert, the difference in prediction is already visible in the first seconds of communication due to the start of the call, then the



Fig. 4 Created alert for the beginning of communication and turning on the webcam

linear method from the family of exponential smoothing methods allows dynamic analysis of network traffic. This method is adaptive to current conditions. Technological solutions for network monitoring and data processing were analyzed and compared. Based on the performed analysis implementation of own platform for network traffic monitoring was created. It uses tshark to capture packets, Kafka to stream them as real-time events, ClickHouse database to store captured packets, InfluxDB to save created forecasts, PostgreSQL database to store monitored configurations and created alerts, application written in Java to perform analysis and create forecasts and a frontend application created with React to provide an intuitive graphical interface for creating configurations and examine created alerts. In the end, the functionality of this experimental implementation was successfully demonstrated on network data with a big amount of video conferencing communication.

VIII. ACKNOWLEDGMENT

This publication has been published with the support of the Operational Program Integrated Infrastructure within project: Research in SANET network and possibilities of its further use and development. ITMS code: 313011W988, co-financed by the ERDF.

REFERENCES

- [1] V. E. Paxson, *Measurements and analysis of end-to-end Internet dynamics*. University of California, Berkeley, 1997.
- [2] M. Czaplewski, "Communication networks as the basis for functioning of the internet," *Procedia Computer Science*, vol. 192, pp. 1770–1778, 2021.
- [3] R. Caceres, *Measurements of wide area internet traffic*. Citeseer, 1989.
- [4] C. Sanders, *Practical packet analysis: Using Wireshark to solve real-world network problems*. No Starch Press, 2017.
- [5] M. H. Bhuyan, D. K. Bhattacharyya, and J. K. Kalita, "Network anomaly detection: methods, systems and tools," *Ieee communications surveys & tutorials*, vol. 16, no. 1, pp. 303–336, 2013.
- [6] P. Barford and D. Plonka, "Characteristics of network traffic flow anomalies," in *Proceedings of the 1st ACM SIGCOMM Workshop on Internet Measurement*, 2001, pp. 69–73.
- [7] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM computing surveys (CSUR)*, vol. 41, no. 3, pp. 1–58, 2009.
- [8] J. D. Hamilton, *Time series analysis*. Princeton university press, 2020.
- [9] I. Svetunkov, "Forecasting and analytics with adam," OpenForecast, 2022. [Online]. Available: <https://openforecast.org/adam/>
- [10] P. Cortez, M. Rio, M. Rocha, and P. Sousa, "Internet traffic forecasting using neural networks," 07 2006, pp. 2635–2642.
- [11] R. J. Hyndman and G. Athanasopoulos, *Forecasting: principles and practice*. OTexts, 2018.

- [12] L. F. Sikos, "Packet analysis for network forensics: A comprehensive survey," *Forensic Science International: Digital Investigation*, vol. 32, p. 200892, 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1742287619302002>
- [13] R. Gerardi, "An introduction to using tcpdump at the linux command line," 2020. [Online]. Available: <https://opensource.com/article/18/10/introduction-tcpdump>
- [14] R. Eddings, "Learning packet analysis with data science," 2018. [Online]. Available: <https://medium.com/hackervalleystudio/learning-packet-analysis-with-data-science-5356a3340d4e>
- [15] M. Tsoukalos, "Using a database to store data capture with tcpdump," vol. 39, no. 1, pp. 35–37, 2014.
- [16] T. Berglund, "Introduction to kafka," 2020. [Online]. Available: <https://kafka.apache.org/intro>
- [17] H. Gupta, V. Harinarayan, A. Rajaraman, and J. D. Ullman, "Index selection for olap," in *Proceedings 13th International Conference on Data Engineering*. IEEE, 1997, pp. 208–219.
- [18] I. Blinkov, "Analytical databases," 2022. [Online]. Available: <https://clickhouse.com/docs/en/>
- [19] J. Volz, "Prometheus documentation," 2022. [Online]. Available: <https://prometheus.io/docs/introduction/overview/>
- [20] P. Dix, "Why time series matters for metrics, real-time analytics and sensor data," 2022. [Online]. Available: <https://www.influxdata.com/time-series-database/>

Prehliadač n-gramových jazykových modelov

Stanislav ONDÁŠ, Michaela NOVOTNÁ

Katedra elektroniky a multimediálnych telekomunikácií, Fakulta elektrotechniky
a informatiky, Technická univerzita v Košiciach, Slovensko

stanislav.ondas@tuke.sk, michaela.novotna@student.tuke.sk

Abstrakt — Cieľom predloženého článku je predstaviť webovú aplikáciu na reprezentáciu štatistických jazykových modelov a na výpočet pravdepodobnosti slovných sekvencií pomocou NoSQL databázy. Pre reprezentáciu n-gramových jazykových modelov bola použitá stĺpcovo-orientovaná databáza Cassandra. Vytvorená aplikácia umožňuje efektívne a komfortné prehliadanie jazykových modelov a evaluáciu slovných sekvencií (viet) pre rôzne druhy modelov, čo umožňuje možnosti porovnávania a skórovania takýchto sekvencií.

Kľúčové slová — databáza, NoSQL, Cassandra, N-gram, jazykový model

I. ÚVOD

Dnešná doba neustále napreduje a s tým je spojené aj množstvo informácií a dát, ktoré je potrebné uchovávať a spracovať. V tomto prípade nám dobre slúžia databázy. Tradičné relačné databázy nemajú problém zvládnuť štruktúrované dáta. Avšak obrovské množstvo dát prináša so sebou aj množstvo neštruktúrovaných a semi-štruktúrovaných dát. Vzhľadom na to vznikli nové typy databáz, ktoré označujeme ako NoSQL (Not only SQL). Tieto databázy sú optimalizované tak aby zvládli dáta, napr. vo forme rozsiahlych textov, obrázkov alebo videí. NoSQL databázy nie sú náhradami ani nástupcami tradičných relačných databáz, predstavujú však optimálnejšie riešenie pre neštruktúrované dáta a pre systémy, v ktorých dochádza k dynamickým zmenám dát.

Typickými predstaviteľmi typov NoSQL databázových systémov sú databázy typu kľúč-hodnota, stĺpcovo-orientované, dokumentové a grafové databázy. Pre reprezentáciu štatistických jazykových modelov sme v tejto práci zvolili databázový systém Cassandra, ktorý patrí medzi stĺpcovo-orientované databázy a bol pôvodne vytvorený spoločnosťou Facebook [4], [5]. V tomto prípade sú dáta ukladané do stĺpcov, na rozdiel od relačného modelu s ukladaním dát do riadkov. Stĺpcovo-orientované databázy ponúkajú dobrú škálovateľnosť a vysokú dostupnosť bez zníženia výkonu. Cassandra využíva jazyk CQL (Cassandra Query Language), ktorý je podobný jazyku SQL. Dokáže spracovať veľké množstvo údajov a operácií za krátky čas.

Štatistické jazykové modely sú dôležitým prvkom v mnohých oblastiach použitia, ako je rozpoznávanie reči alebo strojový preklad. Využívajú tradičné štatistické techniky, ako napr. modelovanie pomocou n-gramov alebo skryté Markovové modely, na reprezentáciu pravdepodobnosti výskytu slov a sekvencií slov v danom jazyku. V jazykovom modelovaní sa defakto jedná o vytváranie pravdepodobnostných modelov, ktoré sú schopné predpovedať nasledujúce slovo v poradí na základe slov, ktoré mu predchádzajú. Medzi najjednoduchšie jazykové modely patria n-gramové modely. V nich sa využívajú n-gramy, ktoré nám predstavujú postupnosť n slov (n-tíc). Tento jazykový model dokáže predpovedať pravdepodobnosť zvoleného slova/sekvencie slov v rámci ľubovoľnej sekvencie slov.

Cieľom predloženej práce je priniesť online nástroj na vizualizáciu n-gramových jazykových modelov s možnosťou výpočtu pravdepodobností analyzovaných sekvencií textu a ich porovnaním naprieč viacerými modelmi. Pre vytvorenie takéhoto nástroja bolo nevyhnutné sa zaoberať transformovaním n-gramových jazykových modelov a ich importovaním do databázového systému, spracovaním vstupných textových reťazcov, vyhľadávaním n-gramových sekvencií a výpočtom pravdepodobnosti výskytu vstupných reťazcov.

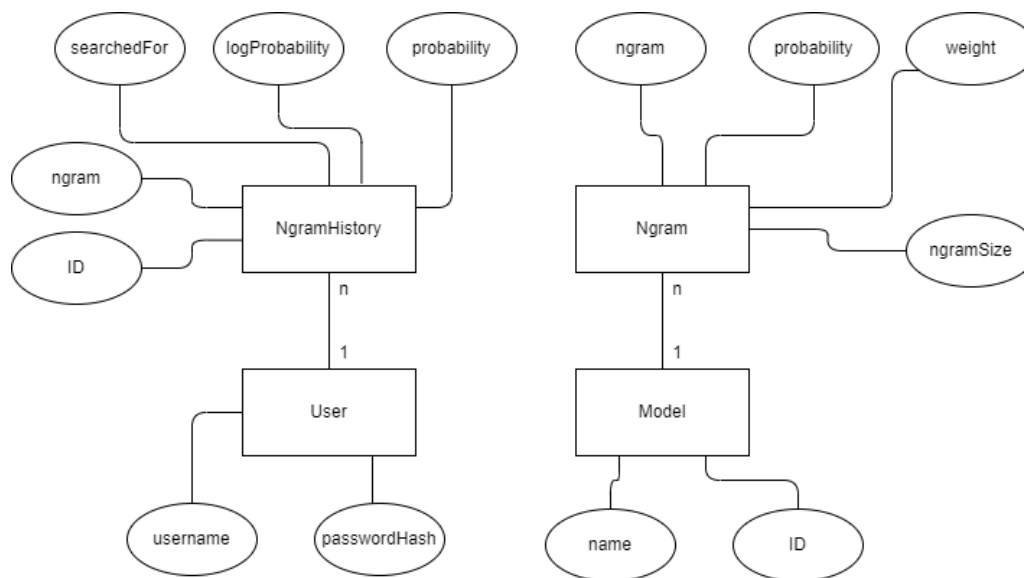
Práca je členená nasledovne. Kapitola II. opisuje návrh aplikácie. Kapitola III. sa venuje implementačným detailom. V kapitole IV. Predstavujeme finálne riešenie webovej aplikácie.

II. NÁVRH DÁTOVÉHO MODELU PRE REPREZENTÁCIU N-GRAMOVÝCH JAZYKOVÝCH MODELOV

Návrh aplikácie pre reprezentáciu štatistických jazykových modelov pozostáva z dvoch hlavných úloh – z návrhu databázy a z návrhu webovej aplikácie nad touto databázou. Pre

reprezentáciu n-gramových modelov sme zvolili databázový systém Cassandra, ktorý je implementáciou stĺpcovo-orientovaného DB systému.

Návrh dátového modelu začal definovaním doménových entít a diagramu dátového modelu, ktorý je zobrazený na Obr.2. Základné identifikované doménové entity sú Ngram, NgramHistory, User a Model.



Obr. 1 Dátový model reprezentácie n-gramového jazykového modelu

Doménová entita Model slúži na identifikáciu modelu. Jej funkcia je skôr formálna a zjednodušuje nám manipuláciu s jednotlivými modelmi. Model môže obsahovať veľké množstvo n-gramov, od niekoľkých stoviek až po niekoľko miliónov.

Doménová entita Ngram umožňuje ukladať jednotlivé n-gramy. Pre uľahčenie vyhľadávania, tabuľka obsahuje referenciu na tabuľku s modelmi. Ďalším dôležitým prvkom je veľkosť n-gramu (ngramSize), čo nám umožňuje definovať počet slov, ktoré tvoria jednotlivé typy n-gramov (unigram – jedno slovo, bigram – dve slová, trigram – tri slová, atď). Ďalšími stĺpcami sú váha (weight) a pravdepodobnosť (probability), ktoré reprezentujú základné parametre jednotlivých slovných n-tíc a sú dôležité pri výpočte výsledných pravdepodobností zadaných slovných postupností.

Doménová entita User reprezentuje používateľa systému. Vzhľadom k tomu, že je potrebný len na autentifikáciu a autorizáciu kvôli spojitosti s históriou vyhľadávania, tak obsahuje len používateľské meno (username) a hash hesla (passwordHash).

Doménová entita NgramHistory nám reprezentuje históriu vyhľadávania N-gramov konkrétnym používateľom. Okrem už vypočítaných pravdepodobností obsahuje aj dátum, kedy bola daná veta vyhľadávaná.

Celkovo v návrhu existujú prepojenia na úrovni 1:N medzi doménovými entitami Model a Ngram. Taktiež medzi entitami User a NgramHistory.

Takto vytvorený dátový model umožňuje plne reprezentovať akékoľvek n-gramové štatistické jazykové modely (ARPA backoff N-gram modely) pomocou stĺpcovo-orientovanej NoSQL databázy.

A. Importovanie modelov do databázy

Ďalším krokom po vytvorení dátového modelu bolo importovanie jazykového modelu do databázy. Z hľadiska možností máme k dispozícii nasledujúce typy prístupov, ktoré pre import môžeme aplikovať:

- manuálny prístup,
- poloautomatizovaný prístup,
- automatizovaný prístup.

Rozhodli sme sa pre automatizovaný prístup, ktorý umožní úplne automatizovať import ARPA modelov do databázy. Pre potreby importovania modelu sme využili voľne dostupný nástroj DS Bulk [1], nakoľko je pri tomto procese dôležitá optimalizácia rýchlosti. DS Bulk umožňuje importovanie dát z CSV formátu. Z týchto dôvodov bolo nutné transformovať vstupné dáta z ARPA do CSV formátu. Pre tento účel sme vytvorili skript v programovacom jazyku Python.

Týmto krokom umožníme automatizáciu aj na iných úrovniach ako napríklad pri spustení kontajnerov, ak by aplikácia bola nasadená v kontajneri popri prípade aj poloautomatizovaný prístup. Rozhodli sme sa spustiť skript po vytvorení inštancií služieb a kontrolérov v aplikácii. To nám zaistí to, že aplikácia bude pripravená až vtedy, keď bude dokončený import všetkých dát. Nevýhodou tohto riešenia je ale dlhšie spúšťanie aplikácie pri importe dát.

III. WEBOVÉ ROZHRAŇIE PREHLIADAČA JAZYKOVÝCH MODELOV

Hlavnou funkcionalitou webového prehliadača jazykových modelov je skórovanie vstupných sekvencií slov pomocou jazykových modelov dostupných v prehliadači. Pod pojmom skórovanie máme na mysli výpočet pravdepodobnosti danej sekvencie slov.

Hlavná stránka webového rozhrania obsahuje vstupné pole na vloženie vstupnej sekvencie slov. Pod vstupným poľom sa nachádza možnosť výberu modelu a história vyhľadávania, ktorá zobrazuje naposledy zadané vstupy v poli.

Po odoslaní vstupného textu sa zobrazí stránka s výpočtom pravdepodobnosti. Hlavným výstupom je logaritmická pravdepodobnosť danej vstupnej sekvencie pre analyzovaný model. Ďalšími zobrazenými informáciami je tabuľka, ktorá obsahuje výpis pravdepodobnosti jednotlivých n-gramov zodpovedajúcich vstupnej vete. (podľa najvyššieho n-gramu v modeli).

Ďalšou funkcionalitou stránky je celkový výpis histórie, ktorá bola vyhľadávaná používateľom.

Vstup do aplikácie je podmienený prihlásením, resp. registráciou používateľa. Po úspešnom prihlásení bude možné využiť všetky možnosti aplikácie.

A. Výpočet pravdepodobnosti vstupnej sekvencie slov

Pre komparatívne účely a pre odvodenie algoritmu výpočtu pravdepodobnosti výskytu vstupnej sekvencie slov sme použili nástroj SRILM toolkit [3].

```
cez cestu prebehla čierna mačka
p( cez | <s> ) = [2gram] 0.000140853 [ -3.851234 ]
p( cestu | cez ... ) = [2gram] 0.006808003 [ -2.16698 ]
p( prebehla | cestu ... ) = [1gram] 6.024388e-07 [ -6.220087 ]
p( čierna | prebehla ... ) = [1gram] 9.540463e-06 [ -5.020431 ]
p( mačka | čierna ... ) = [2gram] 0.005702103 [ -2.243965 ]
p( </s> | mačka ... ) = [2gram] 0.06096674 [ -1.214907 ]
1 sentences, 5 words, 0 00Vs
0 zeroprobs, logprob= -20.7176 ppl= 2837.488 ppl1= 13916.2

file text_dp.txt: 1 sentences, 5 words, 0 00Vs
0 zeroprobs, logprob= -20.7176 ppl= 2837.488 ppl1= 13916.2
```

Obr. 2 Výpočet pravdepodobnosti v nástroji SRILM

Výpočet pravdepodobnosti v systéme SRILM je ilustrovaný na Obr. 2. Pri výpočte sa SRILM toolkit snaží vety rozdeliť na n-gramy najvyššieho rádu, ktorý sa v modeli nachádza. V tomto prípade na n-ticu rádu 3, teda na 3-gramy. Dôležitým prvkom výpočtu je dopočítavanie chýbajúcich n-gramov, ktoré sa uplatní v prípade, ak sa daná n-tica v modeli nenachádza. Tento postup je znázornený na Obr. 3.

```
p(w3|wd1,wd2)= if(trigram exists) p_3(wd1,wd2,wd3)
                else if(bigram w1,w2 exists) bo_wt_2(w1,w2)*p(wd3|wd2)
                else p(wd3|w2)

p(wd2|wd1)= if(bigram exists) p_2(wd1,wd2)
             else bo_wt_1(wd1)*p_1(wd2)
```

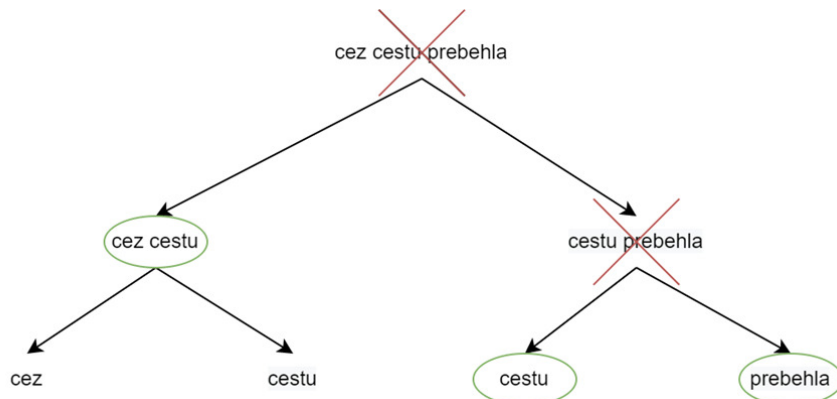
Obr. 3 Postup výpočtu chýbajúceho N-gramu

Aby sme vedeli realizovať výpočet pravdepodobnosti vstupného reťazca musíme vykonať nasledujúce kroky:

- Vyber jazykového modelu
- Vyhľadanie n-gramov, ktoré sa vyskytujú vo vstupnom reťazci a sú tiež obsiahnuté v jazykovom modeli,
- Výpočet pravdepodobnosti a prevod na logaritmickú pravdepodobnosť.

Posledným podporným prvkom je binárny strom tvorený z n-gramov. Ak sa pozrieme na Obr. 2, vidíme, že každé slovo z vety je zahrnuté vo výpočte pravdepodobnosti. Na Obr. 5 je znázornená demonštrácia binárneho stromu, ktorý obsahuje všetky rozložené N-gramy od n-gramu najvyššej úrovne, teda 3-gramu „*cez cestu prebehla*“. Pre výpočet pravdepodobnosti je potrebné, aby sme určili, ktoré n-gramy sa nachádzajú v modeli a ktoré môžeme použiť na

výpočet. Pokiaľ sa daný n-gram nenachádza v modeli, tak z vrcholu ľavého pod-stromu vezmeme váhu n-gramu o stupeň nižšieho rádu, ak sa taký v modeli nachádza. V pravom pod-strome hľadáme, či sa nachádza jeho vrchol v modeli. Ak sa nachádza, použijeme pravdepodobnosť z neho pri výpočte. Ak nie, opakujeme tento proces až kým nedosiahneme najnižší pravý pod-strom, ktorý nemá ďalšie prvky. V prípade, ak sa n-gram najvyššej úrovne nachádza v modeli, berieme do úvahy len jeho pravdepodobnosť.



Obr. 4 Príklad binárneho stromu výpočtu pravdepodobnosti vstupnej sekvencie

Výpočet pravdepodobnosti sa aplikuje na zadaný vstupný text podľa vybraného modelu. Na Obr. 7 sa v hornej tabuľke nachádza text, ktorý bol daný na vstupe, celková logaritická pravdepodobnosť a jeho pravdepodobnosť v percentách. V spodnej tabuľke sa nachádza zápis, ktorý je podobný zápisu v nástroji SRILM. Prvý stĺpec reprezentuje rozdelenie textu na najvyšší N-gram, v tomto prípade 3-gram. Druhý stĺpec nám udáva akú veľkosť N-gramu našiel v modeli. V posledných dvoch stĺpcoch je vypísaná logaritická pravdepodobnosť a pravdepodobnosť v percentách daných N-gramov. Po ich sčítaní dostaneme výslednú pravdepodobnosť.

Výpočet pravdepodobnosti			
Analyzovaný text	Log pravdepodobnosť		Pravdepodobnosť [%]
cez cestu prebehla čierna mačka	-20,71760		1,916e-19
Nájdene ngramy	Veľkosť N-gram	Log pravdepodobnosť	Pravdepodobnosť [%]
P(cez <s>)	2	-3,85123	1,409e-02
P(cestu <s> cez)	2	-2,16698	6,808e-01
P(prebehla cez cestu)	1	-6,22009	6,024e-05
P(čierna cestu prebehla)	1	-5,02043	9,540e-04
P(mačka prebehla čierna)	2	-2,24397	5,702e-01
P(</s> čierna mačka)	2	-1,21491	6,097e+00

Obr. 5 Výpis výpočtu pravdepodobnosti

V histórii nájdeme celkový zoznam zadaných vstupov. Prehľadávanie histórie je možné buď pomocou tlačidla *Zobrazit' viac* alebo pomocou vstupného poľa nad tabuľkou. V ňom je potrebné odlišovať malé a veľké písmená, ktoré boli zadané na vstupe. V prípade, že sme na vstupe zadali slovo s veľkým začiatčným písmenom a vo vyhľadávani ho napíšeme malým začiatčným písmenom, tak nám ho nenájde. V tabuľke vidíme text, ktorý bol daný na vstupe. Vedľa neho je *Detail*. Pomocou neho si dokážeme zobrazit' výpočet pravdepodobnosti bez toho aby sme museli text zadávať na vstupe. Potom nasleduje stĺpec s názvom modelu, ktorý bol vybraný. Predposledné dva stĺpce zobrazujú vypočítanú pravdepodobnosť v logaritme a v percentách. Na konci tabuľky je dátum, kedy bol text zadaný na vstupe.

Hľadaný text	Jazykový model	Log pravdepodobnosť	Pravdepodobnosť [%]	Dátum
cez cestu prebehla čierna mačka	Detail new_apd_20200426_v1_550k-p225e-10	-20,71760	1,916e-19	22:57 12-04-2022
cez cestu prebehla čierna mačka	Detail new_apd_20200426_v1_550k-p225e-10	-20,71760	1,916e-19	22:52 12-04-2022
mačky sú domestikovaný druh	Detail apvv_20180710_v3_550k_p325e-10	-20,60874	2,462e-19	22:52 12-04-2022
cez cestu prebehla čierna mačka	Detail apvv_20180710_v3_550k_p325e-10	-12,25353	5,578e-11	22:52 12-04-2022
cez cestu prebehla čierna mačka	Detail apvv_20180710_v3_550k_p015e-09	-9,43730	3,653e-08	22:52 12-04-2022

Obr. 6 História zadaných vstupov

A. Kontajnerizácia aplikácie

Pre jednoduchšie nasadenie aplikácie sme sa ju rozhodli kontajnerizovať a pripraviť na nasadenie spolu s Cassandra. Webová aplikácia, ktorú sme vytvorili je postavená na Java a frameworku *Spring boot*. Preto je potrebné, aby sme použili *docker image* postavený na Java. Môžeme celkovo rozdeliť build image-u na nasledujúce fázy:

- 1) Build fáza – pozostáva z kompilovania aplikácie a vytvorenia spustiteľného súboru,
- 2) Fáza kopírovania a prípravy na nasadenie – pozostáva z kopírovania spustiteľného súboru, špecifikácie portov, ktoré chceme aby boli dostupné z vonkajšieho prostredia a výbere profilu pri spustení aplikácie.

Pripravili sme štartovací *bash skript*, ktorý sa pripojí na Cassandra a vytvorí požadovaný *keyspace*. Kvôli tomu, ako je postavený oficiálny *docker image* pre Cassandra je na konci tohto skriptu spúšťanie pôvodného spúšťacieho skriptu, ktorý obsahuje Cassandra. Vo finálnej verzii súboru s *docker-compose* definíciami sme ešte pridali prepojenie na jazykové modely, ktoré sú následne automatizovane importované do databázy.

IV. ZÁVER

Predložený článok prezentuje webovú aplikáciu pre prehliadanie štatistických jazykových modelov na báze n-gramov. Pre ich reprezentáciu využíva NoSQL databázový systém Cassandra. Bola navrhnutá reprezentácia n-gramových jazykových modelov v databázovom systéme Cassandra. Toto riešenie prináša možnosť vysokej škálovateľnosti.

Hlavnou funkciou prehliadača je možnosť vyhodnotiť pravdepodobnosť vstupnej sekvencie slov (vety) v danom modeli, resp. vo viacerých modeloch, ktoré boli do systému vložené. Jedným z možných aplikačných použití je analýza vstupného textu na pragmatickej úrovni, kde by mohli byť jazykové modely použité k vyhodnocovaniu správnosti/nesprávnosti tvrdení na základe ich pravdepodobností. Druhým využitím je klasifikácia viet podľa žánru, za predpokladu, že budú v prehliadači vložené doménovo-orientované modely. Potom, je možné klasifikovať vstupné vety do konkrétnej domény podľa najvyššie dosiahnutej úrovne pravdepodobnosti. Ďalšou možnou aplikačnou oblasťou je identifikácia jazyka [6].

POĎAKOVANIE

Táto práca bola podporená Agentúrou na podporu výskumu a vývoja vďaka projektu APVV SK-TW-21-0002, Vedeckou grantovou agentúrou realizáciou výskumného projektu VEGA 1/0753/20 „Robustné rečové technológie metódami hlbokého učenia“ a VEGA 2/0165/21 „Technológie automatického spracovania reči na pomoc v krízových situáciách“ financovaných z prostriedkov Ministerstva školstva, vedy, výskumu a športu SR.

LITERATÚRA

- [1] Docs.datastax.com. 2022. dsbulk :: DataStax Bulk Loader for Apache Cassandra Documentation. [cit. 2022-04-28]. [online] Dostupné na internete: <<https://docs.datastax.com/en/dsbulk/doc/dsbulk/reference/dsbulkCmd.html>>.
- [2] Infinite Scroll. 2022. Infinite Scroll. [cit. 2022-04-28]. [online] Dostupné na internete: <<https://infinite-scroll.com/>>.
- [3] Stolcke, Andreas. "SRILM - an extensible language modeling toolkit." INTERSPEECH (2002).
- [4] Kannan Muthukkaruppan. "The Underlying Technology of Messages" (<http://www.facebook.com/notes/facebook-engineering/the-underlying-technology-of-messages/454991608919>).
- [5] James Hamilton (July 12, 2008). "Facebook Releases Cassandra as Open Source" (<http://perspectives.mvdirona.com/2008/07/12/FacebookReleasesCassandraAsOpenSource.aspx>). Retrieved 2009-06-04.
- [6] Tommi Vatanen, Jaakko J. Väyrynen, and Sami Virpioja. 2010. Language Identification of Short Text Segments with N-gram Models. In Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10), Valletta, Malta. European Language Resources Association (ELRA).

DLL knižnica hašovacej funkcie Blake3 pre Python

¹*Patrik ZELENÁK*, ²*Miloš DRUTAROVSKÝ*

^{1,2}Katedra elektroniky a multimediálnych telekomunikácií, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

¹patrik.zelenak@student.tuke.sk, ²milos.drutarovsky@tuke.sk

Abstrakt – Článok opisuje nový stavebný kryptografický blok, ktorým je vysoko-paralelizovateľná hašovacia funkcia Blake3. Pozornosť je venovaná opisu vnútornej štruktúry funkcie Blake3, ktorá využíva štruktúru Merklóvho hašovacieho stromu. Jadro funkcie Blake3 podporuje spracovanie dát s využitím vektorizovaných SIMD inštrukcií. Článok sa zaoberá prepojením optimalizovanej implementácie funkcie Blake3, napísanej v jazyku Rust, s jazykom Python vo forme DLL knižnice. Uvedené sú výsledky rýchlostných testov, na základe ktorých sú porovnané vybrané implementácie funkcie Blake3. Blake3 je porovnaná s klasickými hašovacími funkciami ako SHA2 alebo SHA3 na systéme s viacerými vláknami.

Kľúčové slová – Kryptografia, Python, Blake3, SIMD, Rust, Dynamická knižnica

I. ÚVOD

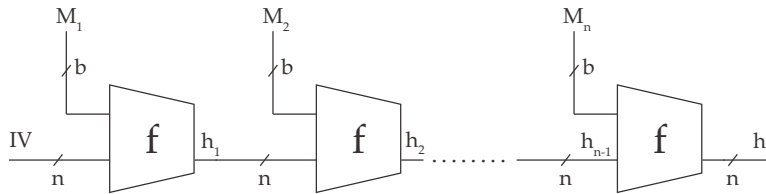
Kryptografia je veda, ktorá využíva matematické metódy utajovania obsahu a preukázateľnosti pôvodu prenášaných správ. Jej cieľom je utajiť obsah prenášaných správ, čo zahŕňa riešenie základných bezpečnostných otázok ako je dôvernosť a integrita dát, autentizácia a autorizácia užívateľov a podobne. Bezpečný prenos a utajenie prenášaného obsahu je možné dosiahnuť s využitím kryptografických algoritmov a protokolov (digitálne podpisy, výmena kľúčov, autentizácia správ, a pod.), ktorých základ tvoria kryptografické primitíva.

V súčasnosti sa pri vývoji softvérových aplikácií využívajú rôzne programovacie jazyky. Zaujímavou voľbou pre vývoj aplikácií je aj jazyk Python [1]. Python je v súčasnosti jedným z najpopulárnejších programovacích jazykov, ktorého využitie stúplo za posledných 5 rokov až o 12,1% v porovnaní s inými programovacími jazykmi. Vďaka svojim výhodám, ako napríklad intuitívna syntax, jednoduchá interakcia s inými programovacími jazykmi, či široká developerská komunita, získal svoje postavenie v oblastiach dátovej vedy (data science), strojového učenia, vývoja webových aplikácií a v mnohých ďalších odvetviach. Programovací jazyk Python disponuje kryptografickými knižnicami ako napríklad Cryptography, PyCryptodome, Hashlib a mnoho ďalších. Opis týchto základných kryptografických knižníc pre jazyk Python je možné nájsť v bakalárskej práci [2]. Kryptografické knižnice sú v práci opisane nie len z hľadiska rozsahu podpory kryptografických algoritmov a protokolov, ale aj z hľadiska rýchlosti spracovania kryptografických algoritmov. Jazyk Python obsahuje aj niekoľko zásadných nevýhod ako napríklad vysoká spotreba operačnej pamäte alebo jeho nízka rýchlosť [3]. Štandardne je vykonávanie programu v jazyku Python pomalšie ako v jazyku C, čo predstavuje jednu z jeho nevýhod. Tieto nevýhody však vieme čiastočne eliminovať prepojením externých modulov, napísaných v inom rýchlejšom programovacom jazyku.

Tento článok obsahuje stručný opis novej paralelizovateľnej hašovacej funkcie Blake3, ktorá využíva štruktúru binárneho Merklóvho stromu [4]. Blake3 je špeciálne navrhnutá pre rýchle hašovanie dát a overenie integrity dát. Rýchlosť funkcie Blake3 značne ovplyvňuje jej štruktúra. Vďaka Merklóvemu stromu je Blake3 vysoko-paralelizovateľná a každé vlákno CPU jednotky môže spracovávať svoj blok dát nezávisle. Jadro funkcie Blake3 je špeciálne navrhnuté pre podporu vektorizovaných SIMD inštrukcií, čo má za následok ďalšie zvýšenie rýchlosti pri spracovaní dát.

II. PARALELIZOVATELNÁ HAŠOVACIA FUNKCIA BLAKE3

Konvenčné hašovacie funkcie (napríklad SHA2 alebo SHA3) spracovávajú dáta sekvenčným spôsobom, to znamená, že sa vstupné dáta typicky rozdelia na bloky dát o určitej veľkosti a tie sa spracovávajú s istou vzájomnou závislosťou (viď. Obr. 1). Každý z blokov je spracovaný kompresnou funkciou f . Výstup kompresnej funkcie sa stáva vstupom pre nasledujúcu kompresnú funkciu. To znamená, že pre spracovanie i -teho bloku M_i , musíme spracovať všetky predchádzajúce bloky dát. Hašovací kód h_n , teda výstup z hašovacej funkcie, dostávame po aplikácii kompresnej funkcie na posledný blok dát M_n .



Obr. 1 Štruktúra konvenčných hašovacích funkcií

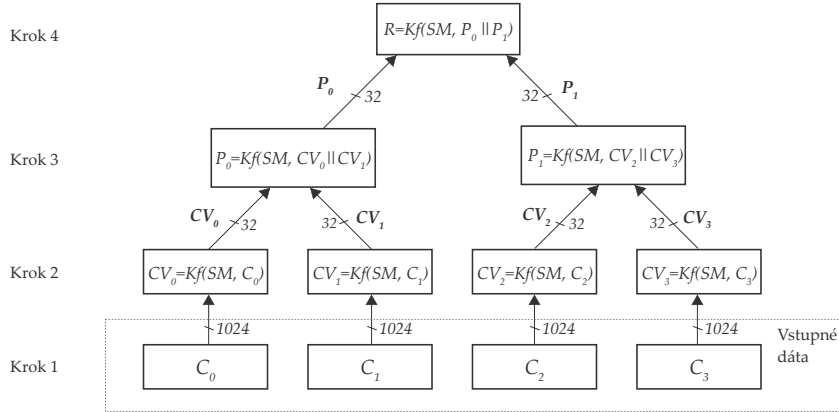
Hašovacia funkcia Blake3 však pracuje úplne inak. Blake3 je moderná hašovacia funkcia vyvinutá v roku 2020 skupinou vývojárov za účelom čo najefektívnejšie a najrýchlejšie hašovať dáta a overiť ich integritu. Blake3 nevyužíva sekvenčné spracovanie dát ako to je u konvenčných hašovacích funkcií ale svoju štruktúru buduje na základoch Merklóvho stromu [5]. Merklóva stromová konštrukcia prináša niekoľko výhod ako napríklad neobmedzenú možnosť paralelizmu pri spracovaní veľkého bloku dát, overovanie integrity streamovaných dát, či overenie integrity dát pri ich sťahovaní z viacerých zdrojov simultánne. K rýchlosti funkcie Blake3 prispieva aj podpora vektorizovaných SIMD inštrukcií, ktorými je možné spracovanie dát výrazne urýchliť na súčasných procesorových platformách.

III. VNÚTORNÁ ŠTRUKTÚRA FUNKCIE BLAKE3

Merklov hašovací strom je hašovacia konštrukcia vynájdená v roku 1982 Ralphom Merklom [6], ktorá vykazuje výhodne vlastnosti pre efektívny paralelizovateľný výpočet integrity dát. Ide o binárnu stromovú štruktúru, ktorej listy reprezentujú bloky vstupných dát C_i s veľkosťou s a každý rodičovský vrchol je odvodený od jeho dvoch potomkov. Realizáciu Merklóvho stromu pre hašovaciu funkciu Blake3 možno opísať v štyroch krokoch.

- 1) **Rozdelenie vstupných dát do blokov C_i .** Vstupné dáta rozdelíme do blokov dát C_i o veľkosti $s = 1024$ bajtov. Tieto bloky dát budeme označovať ako segmenty. Segmenty tvoria listy binárneho hašovacieho stromu.
- 2) **Nezávislé hašovanie blokov dát C_i .** Na jednotlivých segmentoch C_i sa realizuje nezávisle spracovanie dát. V prípade hašovacej funkcie Blake3 sa segmenty C_i spracovávajú kompresnou funkciou, ktorej výstup je označovaný ako zrefazovacia hodnota (chaining value) v binárnom strome. Veľkosť zrefazovacej hodnoty (CV_i) v Merklóvom strome pre funkciu Blake3 je 32 bajtov. Zrefazovacie hodnoty slúžia na tvorbu nadradených rodičovských uzlov Merklóvho stromu. Práve kvôli nezávislému spracovaniu segmentov poskytuje Merklóva stromová štruktúra neobmedzenú možnosť paralelizmu pri spracovaní dát.
- 3) **Tvorba a hašovanie rodičovských uzlov P .** Každý rodičovský uzol pozostáva práve z dvoch hodnôt a to zo zrefazovacej hodnoty ľavého potomka a zrefazovacej hodnoty pravého potomka. Na Obr. 2 je vidieť, že rodičovský uzol P_0 pozostáva z jeho ľavého potomka C_0 a pravého potomka C_1 , konkrétne z ich zrefazovacích hodnôt CV_0 a CV_1 . Hašovaním rodičovského uzla získavame opäť zrefazovacia hodnotu, ktorá sa stáva jedným zo vstupov svojho nadriadeného rodičovského uzla. Takéto hierarchické spracovávanie uzlov, smerom od listov ku koreňu, vykonávame až kým sa nedostaneme ku koreňu binárneho stromu.
- 4) **Hašovanie koreňa stromu R .** Vo všeobecnosti sa hašovanie koreňa stromu realizuje rovnakým spôsobom ako hašovanie rodičovských uzlov. Na Obr. 2 môžeme vidieť, že koreň stromu R má dvoch potomkov P_0 a P_1 . Zrefazovacie hodnoty potomkov koreňa tvoria vstup pre kompresnú funkciu $Kf()$ a výstupom $Kf()$ získavame hašovací kód Merklóvej hašovacej štruktúry v hašovacej funkcii Blake3. Teda po aplikácii

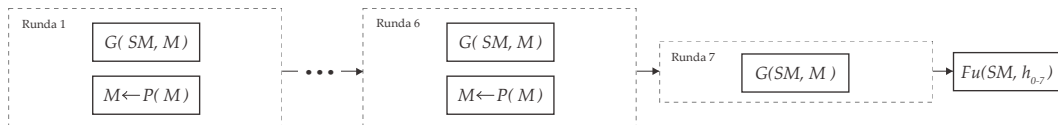
$Kf(SM, P_0 || P_1)$ získavame výstupný hašovaci kód funkcie Blake3.



Obr. 2 Merklova stromová štruktúra hašovacej funkcie Blake3

Obr. 2 demonštruje rozdelenie vstupných dát na segmenty (C_0, C_1, C_2, C_3) o veľkosti 1024 bajtov. Ak je posledný segment menší ako 1024 bajtov, bude mu pridaná vhodná výplňová schéma (padding) v podobe nulových bajtov. Následne sa segmenty spracovávajú funkciou $Kf()$, ktorá na výstup vracia zrefazovaciu hodnotu. Zrefazovacie hodnoty dvoch segmentov tvoria vstupy pre výpočet zretazovacej hodnoty ich nadradeného rodičovského uzla. Kompresná funkcia $Kf()$ požaduje dva vstupné parametre, a to stavovú maticu SM s veľkosťou 64 bajtov a blok vstupných dát M taktiež o veľkosti 64 bajtov. Listy a rodičovské uzly sa spracovávajú odlišným spôsobom. V prípade rodičovských uzlov je hodnota M rovná $CV_l || CV_p$, teda M je rovné spojeniu (concatenation) pravej a ľavej zrefazovacej hodnoty. Pre listy binárneho stromu platí, že sa list (1024 bajtov) rozdelí na 64-bajtové subbloky M_0, \dots, M_{15} , ktoré sa následne spracovávajú sekvenčne [2].

Kompresná funkcia $Kf()$ spracováva dáta v 7 rundách. Každá runda, okrem poslednej, je tvorená tzv. G funkciou a následnou permutáciou vstupných dát, čo možno vidieť na Obr. 3. V poslednej runde sa tiež aplikuje G funkcia, no permutácia vstupných dát sa už nevykonáva. Konečným krokom kompresnej funkcie $Kf()$ je transformácia stavovej matice označená ako $Fu()$.



Obr. 3 Bloková schéma kompresnej funkcie $Kf()$

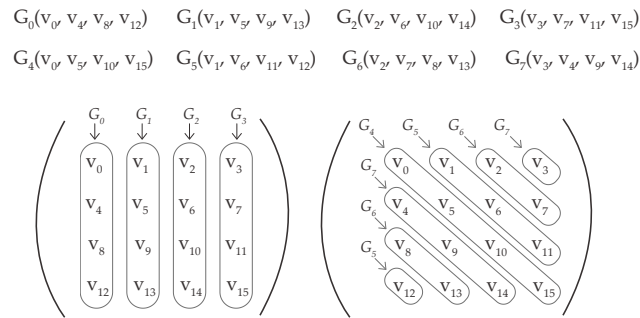
Blake3 podporuje spracovanie dát pomocou vektorizovaných SIMD inštrukcií, ktoré využíva práve G funkcia. G funkcia má za úlohu zmeniť stav matice SM v závislosti od jej aktuálneho stavu a bloku vstupných dát. Stavovú maticu SM (64 bajtov) možno zapísať v podobe šestnástich 32-bitových slov. Stavová matica je inicializovaná pri každom spracovaní bloku dát a obsahuje parametre ako inicializačný vektor IV , zrefazovacie hodnoty h_{0-7} , počítadlo segmentov t_{0-1} a pod. [7]. Tieto slová možno vyjadriť v matici 4×4 nasledovne:

$$\begin{pmatrix} h_0 & h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 & h_7 \\ IV_0 & IV_1 & IV_2 & IV_3 \\ t_0 & t_1 & b & d \end{pmatrix} \rightarrow \begin{pmatrix} v_0 & v_1 & v_2 & v_3 \\ v_4 & v_5 & v_6 & v_7 \\ v_8 & v_9 & v_{10} & v_{11} \\ v_{12} & v_{13} & v_{14} & v_{15} \end{pmatrix}$$

Obr. 4 Stavová matica SM 4×4 , ktorá obsahuje šestnásť 32-bitových slov vyjadrených aliasmi v_{0-15}

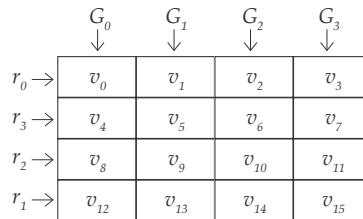
Matica vľavo (viď. Obr. 4) predstavuje parametre stavovej matice SM a matica vpravo predstavuje aliasy hodnôt stavovej matice (matice vľavo), ktoré slúžia na jednotný opis parametrov stavovej matice [7].

G funkcia pozostáva z 8 operácií G_{0-7} a je aplikovaná na každý stĺpec matice 4×4 , a potom aj na každú diagonálu. Spracovanie stĺpcov a diagonál prebieha v 8 operáciách G_{0-7} paralelne. Na Obr. 5 môžeme vidieť grafické znázornenie a tiež matematické vyjadrenie týchto G operácií, pričom slová v_{0-15} predstavujú aliasy stavovej matice SM .



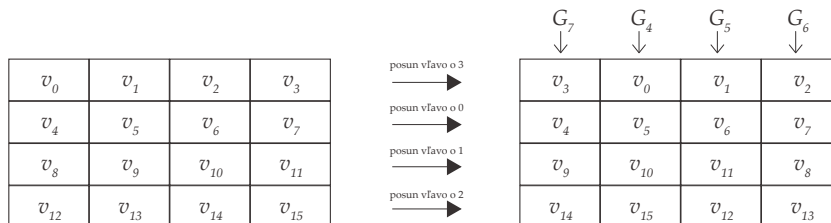
Obr. 5 Grafické znázornenie a matematické vyjadrenie realizácie G operácií paralelným spôsobom

Funkcia Blake3 podporuje 3 rôzne spôsoby využitia SIMD inštrukcií. Článok opisuje prvý z troch spôsobov využitia SIMD inštrukcií, ktorý je podobný ako u predchodcov funkcie Blake3 (Blake2b, resp. Blake2s). Slová matice sa rozdelia po riadkoch do štyroch 128-bitových vektorov r_{0-3} (Obr. 6). Následne sa vykoná vektorizovaná G funkcia (ktorá pracuje s vektorovými inštrukciami) na všetkých 4 stĺpoch paralelne. To znamená, že operácie G_{0-3} (spracovanie stĺpcov) sa vykonávajú naraz po riadkoch pomocou SIMD inštrukcií. Inak povedané, každý riadok obsahuje prvé slovo z jednotlivých G_{0-3} operácií a spracovaním prvého riadku vykonáme spracovanie prvého slova vo všetkých G_{0-3} operáciách. Takýmto spôsobom pokračujeme v spracovaní všetkých vektorov r_{0-3} .



Obr. 6 Paralelné spracovanie stĺpcov operáciami G_{0-3}

Po spracovaní stĺpcov (vykonanie operácií G_{0-3}) nasleduje diagonalizácia riadkov (rotácia riadkov), čo zabezpečí vhodné preusporiadanie riadkov tak, aby bolo možné opäť vykonať vektorizované spracovanie stĺpcov, tentokrát pre operácie G_{4-7} (viď. Obr. 7). Po vykonaní operácií G_{4-7} , nasleduje inverzná diagonalizácia riadkov, teda preusporiadanie riadkov do pôvodného stavu. Takéto využitie SIMD inštrukcií je vhodné aplikovať pri spracovávaní malého množstva vstupných dát. V prípade veľkého množstva vstupných dát je vhodné využiť niektorý z ďalších dvoch spôsobov využitia SIMD inštrukcií [7].



Obr. 7 Rotácia vektorov a následné paralelné spracovanie stĺpcov operáciami G_{4-7} operácií

IV. PARALELIZOVATELNÁ DYNAMICKÁ KNIŽNICA BLAKE3 PRE JAZYK PYTHON

Autori hašovacej funkcie Blake3 napísali optimalizovanú, plne paralelizovateľnú implementáciu funkcie Blake3 v jazyku Rust [8]. Implementácia funkcie Blake3 je dostupná

aj v jazyku C, avšak nie je paralelizovateľná. Programovací jazyk Rust je relatívne nový programovací jazyk, ktorý bol prvýkrát predstavený v roku 2014 spoločnosťou Microsoft. Ide o programovací jazyk, ktorý by vďaka svojej rýchlosti a bezpečnosti mohol predstavovať alternatívu programovacieho jazyka C resp. C++. Vývojári jazyka Rust propagujú jeho schopnosť pomáhať vývojárom vytvárať rýchle a bezpečné aplikácie. Rust neobsahuje garbage collector, runtime alebo manuálnu správu pamäte, čo súvisí s jeho rýchlosťou a bezpečnosťou a taktiež zabráňuje chybám segmentácie pamäte. Rust je špeciálne navrhnutý pre bezpečnú správu pamäte a vývoj zreťazených (pipeline) a paralelizovateľných aplikácií.

A. Prepojenie optimalizovanej implementácie funkcie Blake3 s jazykom Python

Prepojili sme optimalizovanú implementáciu Blake3, ktorej zdrojový kód je vytvorený v jazyku Rust, s jazykom Python. Prepojenie externého Rust modulu (Blake3) sme realizovali vytvorením dynamickej Python knižnice, ktorú je možné použiť v Python aplikácií na hašovanie súboru s využitím funkcie Blake3. Výhodou tohto prepojenia je zachovanie rýchlosti a využitie paralelizmu, ktorú nám ponúka jazyk Rust a zároveň využitie komfortu jazyka Python.

B. Nástroje pre vytvorenie dynamickej Python knižnice

Pre vytvorenie dynamického Python modulu sme využili nástroje zobrazené v Tabuľke 1.

Tabuľka 1
Nástroje použité na vytvorenie dynamickej knižnice

Programovací jazyk/modul	Verzia
Rust (programovací jazyk)	rustc 1.56.1
Cargo (balíčkový manažér pre Rust)	cargo 1.56.0
Python (programovací jazyk)	Python 3.8.10
pip (balíčkový manažér pre Python)	pip 22.0.3
PyO3 (prepájací modul)	0.15.1
Blake3 (hašovacia funkcia)	1.3.1

Pred samotným vytvorením dynamického Python modulu je nutné namapovať kód externého Rust modulu s jazykom Python. Takéto mapovanie je možné realizovať pomocou Rust balíčka PyO3 [9]. Pri vytváraní dynamického Python modulu sme postupovali manuálnym spôsobom, ktorý sa javí ako najjednoduchší. Manuálne vytvorenie dynamického Python modulu spočíva vo vytvorení DLL knižnice v jazyku Rust. Následou modifikáciou DLL knižnice získame dynamický Python modul. Bližšie detaily je možné nájsť v práci [2].

V. EXPERIMENTÁLNE TESTOVANIE RÝCHLOSTI FUNKCIE BLAKE3

Vytvorili a experimentálne sme otestovali Python implementáciu hašovacej funkcie Blake3, ktorá je založená na prepojení optimalizovanej implementácie napísanej v jazyku Rust.

A. Porovnanie rôznych implemetácií hašovacej funkcie Blake3

Pre vytvorenú paralelizovateľnú DLL knižnicu funkcie Blake3 sme vykonali rýchlostné testy, ktorých výsledok je zobrazený v Tabuľke 2.

Tabuľka 2
Rýchlostný test hašovacej funkcie Blake3, vykonaných na procesore I5-8250U a OS Windows 10, pre rôzne implementácie funkcie Blake3

Implementácia	Vlákná (Threads)/ Jadrá	100 MB	200 MB	1 GB
Opt. Rust impl. (vytvorená autormi)	1 vlákno	0,064 s	0,128 s	0,655 s
	2 vlákna	0,033 s	0,069 s	0,341 s
	3 vlákna	0,023 s	0,043 s	0,242 s
	4 vlákna	0,021 s	0,036 s	0,201 s
	5 vlákien	0,019 s	0,031 s	0,165 s
	6 vlákien	0,015 s	0,029 s	0,153 s
	7 vlákien	0,014 s	0,028 s	0,142 s
	8 vlákien	0,013 s	0,027 s	0,138 s
rust_py_blake3 (využíva vytvorenú DLL knižnicu)	1 vlákno	0,065 s	0,129 s	0,665 s
	2 vlákna	0,035 s	0,089 s	0,346 s
	3 vlákna	0,024 s	0,046 s	0,301 s
	4 vlákna	0,022 s	0,043 s	0,203 s
	5 vlákien	0,021 s	0,034 s	0,171 s
	6 vlákien	0,018 s	0,032 s	0,167 s
	7 vlákien	0,017 s	0,031 s	0,156 s
	8 vlákien	0,016 s	0,029 s	0,146 s

Tabuľka 2 predstavuje výsledky rýchlostného testu funkcie Blake3 pre optimalizovanú implementáciu funkcie Blake3 a DLL implementáciu `rust_py_blake3`, ktorá využíva vytvorenú dynamickú Python knižnicu. V prípade optimalizovanej paralelizovateľnej implementácie napísanej v programovacom jazyku Rust a nami vytvorenej implementácie `rust_py_blake3` bolo možné hašovanie otestovať z hľadiska rýchlosti na rôznom počte aktívnych vlákien procesora. Z meraní vyplýva, že implementácia napísaná v jazyku Rust je v porovnaní s implementáciou `rust_py_blake3` rýchlejšia v priemere o približne 6.7%.

B. Porovnanie funkcie Blake3 s inými modernými hašovacími funkciami

Hašovacia funkcia Blake3, spolu s ďalšími modernými hašovacími funkciami, bola rýchlostne testovaná. Na základe výsledkov meraní sme vybrané hašovacie funkcie porovnali.

Tabuľka 3
Rýchlostný test hašovacej funkcie Blake3, vykonaných na procesore I5-8250U a OS Windows 10, pre rôzne implementácie funkcie Blake3

Porovnanie vybraných hašovacích funkcií			
Súbor	SHA-256	SHA3-256	<code>rust_py_blake3</code>
20 MB	0,0958 s	0,1177 s	0,0051 s
100 MB	0,4807 s	0,5901 s	0,016 s
500 MB	2,3168 s	2,8908 s	0,0696 s
2 GB	10,0687 s	12,0961 s	0,2934 s

Tabuľka 3 zobrazuje porovnanie konvenčných hašovacích funkcií SHA256 a SHA3-256 s vytvorenou DLL knižnicou hašovacej funkcie Blake3. Vybrané konvenčné hašovacie funkcie (z rodiny SHA2 a SHA3) sú implementované v kryptografickej knižnici Cryptography. Z Tabuľky 3 je vidieť, že vytvorená DLL implementácia hašovacej funkcie Blake3, využívajúca všetkých 8 dostupných vlákien, je podľa očakávaní výrazne rýchlejšia ako konvenčné hašovacie funkcie a to niekoľko-násobne.

VI. ZÁVER

Článok sa venuje opisu kryptografickej hašovacej funkcie Blake3, ktorá svojou paralelizovateľnou štruktúrou založenou na Merklovom hašovacom strome vykazuje výhodné vlastnosti pre rýchle a paralelné hašovanie dát. Článok tiež opisuje princíp využitia SIMD inštrukcií vo funkcii Blake3, čo má vplyv na ďalšie zvýšenie rýchlosti hašovania dát. Využitá optimalizovaná implementácia hašovacej funkcie Blake3 je napísaná v jazyku Rust. Článok opisuje prepojenie externého Rust modulu s jazykom Python prostredníctvom dynamickej knižnice. Prepojenie modulu Blake3 s jazykom Python spája výhodne vlastnosti jazyka Rust, ako napríklad rýchlosť vykonávania programu, s komfortom jazyka Python. Blake3 je vhodná na rýchle hašovanie veľkého množstva dát s využitím moderných viac-jadrových procesorov.

VII. POĎAKOVANIE

Tento príspevok vznikol vďaka podpore Vedeckej grantovej agentúry Ministerstva školstva, vedy, výskumu a športu SR podporenej grantom VEGA 1/0584/20.

LITERATÚRA

- [1] "Python language," [online] [cit. 2022-05-17], Dostupné z: <https://www.python.org/>.
- [2] P. Zelenák, "Kryptografia v pythone," [online] [cit. 2022-05-17], košice: Technická univerzita v Košiciach, Fakulta elektrotechniky a informatiky, 2022. 87s.
- [3] W. J. Chun, *Core Python Programming*. USA: Prentice Hall PTR, 2006.
- [4] J. Chapweske, "Tree hash exchange format (thex)," [online] [cit. 2022-05-17], Dostupné z: <https://adcs.sourceforge.io/draft-jchapweske-thex-02.html>.
- [5] M. Szydło, "Merkle tree traversal in log space and time," [online] [cit. 2022-05-17], Dostupné z: https://link.springer.com/content/pdf/10.1007/978-3-540-24676-3_32.pdf.
- [6] R. Merkle, "Method of providing digital signatures," [online] [cit. 2022-05-17], Dostupné z: <https://worldwide.espacenet.com/patent/search/family/022107098/publication/US4309569A?q=pn%3DUS4309569>.
- [7] J. O'Connor, J.-P. Aumasson, S. Neves, and Z. Wilcox-O'Hearn, "Blake3," [online] [cit. 2022-05-17], Dostupné z: <https://github.com/BLAKE3-team/BLAKE3-specs/blob/master/blake3.pdf>.
- [8] "Rust language," [online] [cit. 2022-05-17], Dostupné z: <https://www.rust-lang.org/>.
- [9] "Pyo3," [online] [cit. 2022-05-17], Dostupné z: <https://docs.rs/pyo3/latest/pyo3/>.

Návrh a Riadenie Inteligentného Skleníka

¹Ján KAŇUCH, ²Bc. Ivana SAVINECKÁ

^{1,2} Katedra elektrotechniky a mechatroniky, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Letná 1/9, 04200 Košice-Sever, Slovenská republika

¹jan.kanuch@tuke.sk, ²ivana.savinecka@student.tuke.sk

Abstrakt — V tomto príspevku je prezentovaný návrh a konštrukcia riadeného inteligentného skleníka. V úvode je popísaný účel skleníkov, definuje sa inteligentný skleník, jeho funkcie a výhody v súčasnosti. V ďalšej časti príspevku je popísaný a vypracovaný grafický návrh skleníka a jeho komponentov tak, aby bol riaditeľný. Ďalšia časť sa venuje elektrickej časti skleníka, kde je zobrazená bloková schéma zapojenia riadenia. Následne je popísaná samotná realizácia inteligentného skleníka a prezentovaný je aj program, ktorý riadi celý inteligentný skleník.

Kľúčové slová — skleník, automatizácia, riadenie, Arduino Uno, inteligentný skleník

I. ÚVOD

Vzhľadom na dnešnú klimatickú situáciu, globálne a ekonomické zmeny bude v blízkej budúcnosti nevyhnutné zaradenie inteligentných skleníkov do nášho každodenného života. Jednoduchý záhradný skleník nepotrebuje nutne automatizáciu, ako napríklad počítačom riadené tienidlá, zavlažovanie, vetranie, reguláciu teploty či umelé osvetlenie. Treba si však uvedomiť, že pri dnešných globálnych výkyvoch počasia je pri množstevnom pestovaní automatizácia skleníkov nevyhnutná. Čoraz viac úzko špecializovaných pestovateľských firiem, ale aj poľnohospodárov a domácností využíva pozitíva inteligentných skleníkov. [1]

Pre inteligentné skleníky je k dispozícii mnoho rôznych typov senzorov, ktoré merajú a hlásia údaje z monitorovacieho prostredia a posúvajú dáta do riadiacej automatizácie. Je dôležité si rozmyslieť čo rastlina vyžaduje, a na základe toho zväziť výber senzorov, zariadení aby sa dosiahol optimálny rast a život plodín. Zariadenia by mali byť schopné zvládnuť skleníkové prostredie, ktoré si zvyčajne vyžaduje osobitnú pozornosť s elektrickými zdrojmi a senzormi, ktoré môžu byť citlivé na striekajúcu vodu, slnečné svetlo a vlhkosť. V tabuľke 1 je uvedený prehľad senzorov na riadenie rôznych funkcií inteligentného skleníka.

Tabuľka 1
Prehľad senzorov a riadenia funkcií inteligentného skleníka

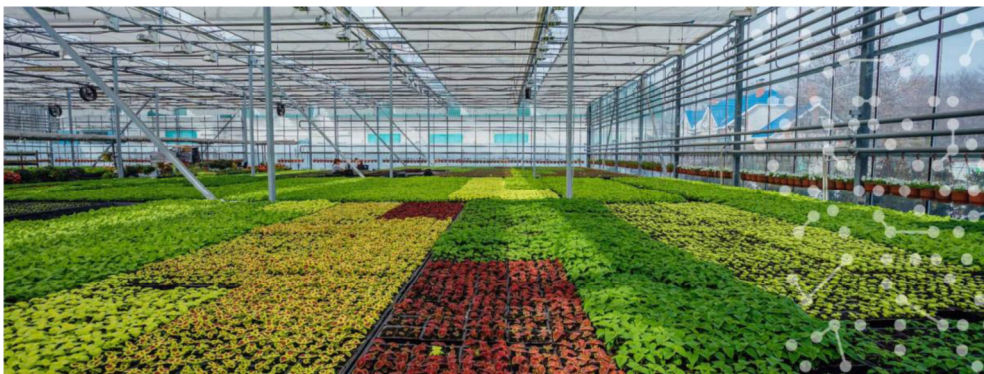
Senzor	Monitorovanie problému	Riešenie problému
Senzor teploty	Nárast teploty nad optimálnu	Zapnutie ventilátorov na bočných stranách skleníka/ otváranie okien
	Nižšia teplota, ako je optimálna	Zapnutie vykurovania v skleníku
Senzor vlhkosti/ pôdnej vlhkosti	Nárast vlhkosti nad optimálnu	Zapnutie ventilátorov na bočných stranách skleníka/ otváranie okien
	Nižšia vlhkosť, ako je optimálna	Zapnutie kvapkového zavlažovania
Svetelné senzory	Veľká intenzita svetla počas dlhšej doby	Zapnutie plynulého vyťahovania tienidla v podobe žalúzií
	Nízka svetelnosť počas dlhšej doby	Zapnutie umelého osvetlenia na dobu určitú
Senzory pôdneho pH	Nameranie zvýšeného pH pôdy Nameranie zníženého pH pôdy	Čerpadlom nasatý neutralizačný roztok pH pôdy do vodnej nádržky.
Senzor kvality vzduchu/ oxidu uhličitého	Nameranie nedostatku oxidu uhličitého	Zapnutie ventilátorov na bočných stranách skleníka/ otváranie okien

Pre pestovateľa je riadenie environmentálnych faktorov vrátane tepla, vlhkosti, úrovne osvetlenia a vetrania veľmi náročné. Z tohoto dôvodu automatizované kontroly niektorých

faktorov pomáhajú zmierniť veľkú časť záťaže a problémov, ktorým čelia pestovatelia. Automatické ovládacie prvky možno plne integrovať do všetkých častí skleníkového systému, aby sa prostredie v skleníku prispôbilo potrebám plodín bez potreby neustáleho nastavovania. Inštalácia automatizačného systému v komerčných skleníkoch pomôže vybudovať konzistentné a predvídateľné cykly, ktoré udržia v skleníku ideálne pestovateľské prostredie [2].

Jednou z najväčších výhod, ktorú môže automatizácia poskytnúť, je schopnosť znížiť celkové výrobné náklady. Integrovaný automatizovaný systém si vyžaduje kapitálové náklady, no dlhodobé úspory nám celkové náklady za energiu znížia. Technológia sa postupom času zlepšila, zvýšila sa dostupnosť pre každého pestovateľa s akýmkoľvek rozpočtom. Automatizačný systém zabezpečí efektívne využitie energie, aby použité systémy fungovali ako súdržná jednotka, čo ušetrí celkové výrobné náklady a náklady za pracovnú silu.

Integrovaný automatizovaný systém v skleníku môže pomôcť pestovateľom na miestach, kde môže byť klíma obmedzujúca pre pestovanie. Inteligentné skleníky (Obr. 1) sa dajú využívať celoročne pri vhodnom nastavení monitorovacích parametrov a zaisťujú nám, že produkt môžeme pestovať bez ohľadu na lokalitu a výkyvy počasia, ktoré sú v dnešnej dobe častejšie ako v minulosti [2].



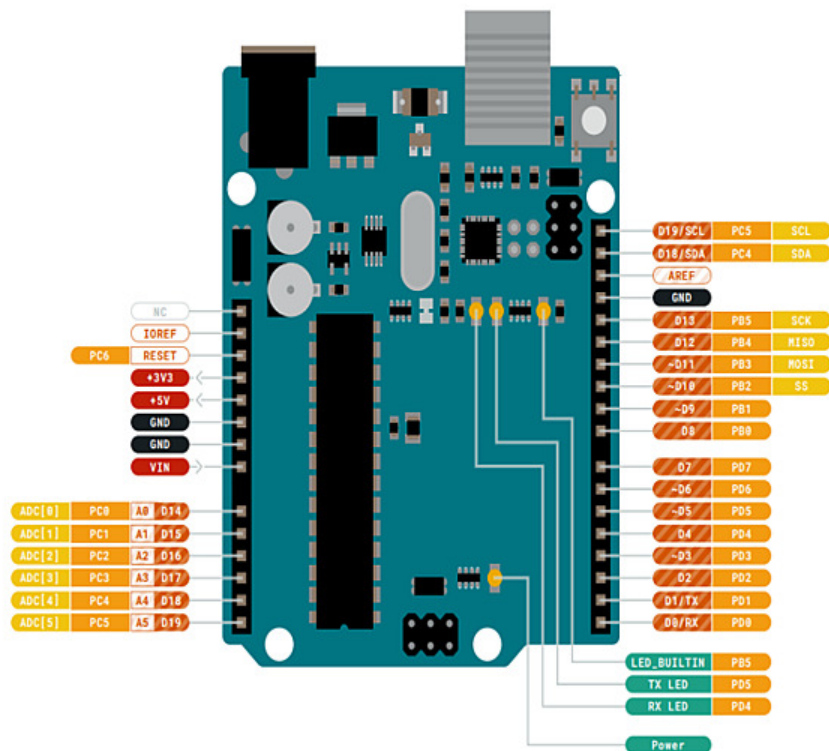
Obr. 1 Pestovanie rôznych druhov rastlín v inteligentnom skleníku [3].

II. KOMPONENTY POTREBNÉ NA AUTOMATIZÁCIU INTELIGENTNÉHO SKLENÍKA

V tejto kapitole sú popísané komponenty, ktoré sú použité pri výstavbe inteligentného skleníka.

Arduino UNO R3 – riadiaca jednotka

UNO R3 je založený na platforme Arduino. Arduino je platforma s otvoreným zdrojovým kódom, ktorá sa používa na vytváranie projektov v oblasti elektroniky. Arduino pozostáva z fyzickej programovateľnej dosky plošných spojov (mikrokontrolér) a časti softvéru alebo IDE (Integrované vývojové prostredie), ktoré beží na počítači a slúži na písanie a nahrávanie počítačového kódu na fyzickú dosku [4], [5].



Obr. 2 Rozmiestnenie pinov na doske Arduino UNO R3 [5].

LM35 senzor- senzor teploty vzduchu

LM35 je procesný snímač teploty s integrovaným obvodom, ktorého výstupné napätie sa mení v závislosti od okolitej teploty [6]. Je malý, lacný, používa sa na vzdialené meranie. Môže byť ľahko prepojený s akoukoľvek vývojovou platformou, ako je aj Arduino.

DHT11 senzor – senzor teploty a vlhkosti vzduchu

Z dôvodu menšieho rozsahu teplôt sa daný senzor využije iba na meranie vlhkosti vzduchu. Je to presný a lacný snímač, ktorý obsahuje ADC na prevod analógových hodnôt vlhkosti a teploty. Na prevod dát sa používa 8-bitový mikrokontrolér v senzore [7].

FC- 28 snímač a modul snímača vlhkosti pôdy

Modul sa využíva na detekciu vlhkosti v pôde. Meria objemový obsah vody v pôde a ako výstup udáva úroveň vlhkosti. Modul má digitálne aj analógové výstupy a potenciometer na nastavenie prahovej úrovne. Dve veľké odkryté sondy sa používajú na prechod prúdu cez pôdu. Čím viac vody je v pôde, tým lepšia bude vodivosť medzi sondami a výsledkom bude nižší odpor a vyššie napätie na výstupe [8].

DS18B20 senzor - senzor teploty pôdy

Na meranie teploty pôdy je použitý vodotesný senzor DS18B20. Pre získanie dát zo senzora sa pripojí digitálny výstup senzora na digitálny vstup Arduina UNO. Ku zdroju napätia pre senzor je nutné zapojiť do série rezistor 4.7 k Ω [9].

GL5528 senzor - svetelný senzor s fotorezistorom

Modul svetelného senzora využíva na detekciu intenzity osvetlenia prostredia fotorezistor. Odpor senzora klesá pri zvyšujúcej sa intenzite svetla [10].

Senzor MH-Z19- senzor na snímanie kvality ovzdušia

Modul pracuje na princípe infračerveného sledovania CO₂, je nezávislý od obsahu O₂ vo vzduchu. Senzor má funkciu samokalibrácie – najnižšiu nameranú hodnotu kontaminácie CO₂ bude považovať za čistý vzduch. Táto funkcia je voliteľná [11].

pH sonda s modulom, senzor na meranie pH vody v pôde

Skratka pH znamená množstvo vodíka v roztoku, pričom pH modul a sonda nám merajú koncentráciu vodíkových iónov vo vode pôdy. Hodnota pH sa pohybuje od 1-14. Ideálna

hodnota pH sa pohybuje v hodnotách 7 – 7.2, kedy hovoríme o neutrálnom pH. Ak sa nameria v roztoku menšie pH ako 7, tak sa roztok považuje za kyslý, naopak, pri nameraní viac ako 7.2 pH považujeme roztok za zásaditý. Modul a sonda sú navzájom prepojené BNC konektorom [12].

Senzor hladiny vody WLS-001

Senzor má sériu desiatich odkrytých medených stôp. Stopy fungujú ako premenlivý odpor a odpor sa mení v závislosti od hladiny vody. Čím viac je snímač ponorený do vody, tým lepšia je vodivosť a nižší je odpor. Na výstupe senzora je výstupné napätie závislé na jeho odpore. Senzor hladiny vody sa využije na meranie hladiny vody v nádrži určenej na zavlažovanie [13].

Topný kábel 12V

Jeden topný kábel sa použije na ohrev pôdy a druhý na vykurovanie vzduchu v skleníku. Tepelné jadro je odporový vodič spletený z viac drôtov. Pri prechode elektrického prúdu vodičom sa začne topný kábel zahrievať. Topný kábel sa zapojí cez relé na 12V pričom signál pre spustenie vykurovania bude prichádzať z Arduina [14].

DC-DC menič napätia 12/5 V

Menič napätia z 12 V DC na 5 V DC sa využije na napájanie ventilátorov a čerpadiel. Menič napätia sa pripojí na transformátor a výstup meniča bude napájať koncové zariadenia pomocou relé [15].

Hodinový modul DS1302

Hodinový modul dokáže počítať sekundy, minúty, hodiny, dni, mesiace a roky. Ráta sa aj s priestupnými rokmi a dátumy sú validné do roku 2100. Na module sa nachádza slot pre mincovú batériu, vďaka ktorej sa zachová čas aj po odpojení klasického napájania. Hodinový modul sa využije pri programovaní podmienok osvetlenia [16].

LCD displej 16x2, I2C

Modrý LCD displej s bielym LED podsvietením a ovládaním I2C je vhodný pre jednoduché projekty. Výhodou je ovládanie zbernice I2C. Displej je prepojený s Arduinom pomocou 4 kolíkov - +5 V, GND, SDA, SCL [17].

Mimo vyššie popísaných komponentov, ktoré sú použité pri výstavbe inteligentného skleníka sú potrebné aj tieto: sieťový zdroj TLP-12-75 (230 V AC/12 V DC) [18], 2 x 2-pinový ventilátor (5 V) [19], malá vodná pumpa (vodné čerpadlo) [20], 8 kanálové relé s jumperom H/L 5V [21] a LED pásik pre osvetlenie skleníka (12 V) [22].

III. NÁVRH ELEKTRICKEJ SCHÉMY INTELIGENTNÉHO SKLENÍKA

Jednotlivé komponenty potrebné na reguláciu skleníka sa vzhľadom na ich inštaláciu rozdelili na dve časti. Hlavná časť je uložená na dne skleníka a z nej sú napájané senzory, ventilátory, čerpadlá a LED pásik, ktoré sú rozmiestnené po skleníku vzhľadom na ich využitie.

A. Návrh blokovej elektrickej schémy

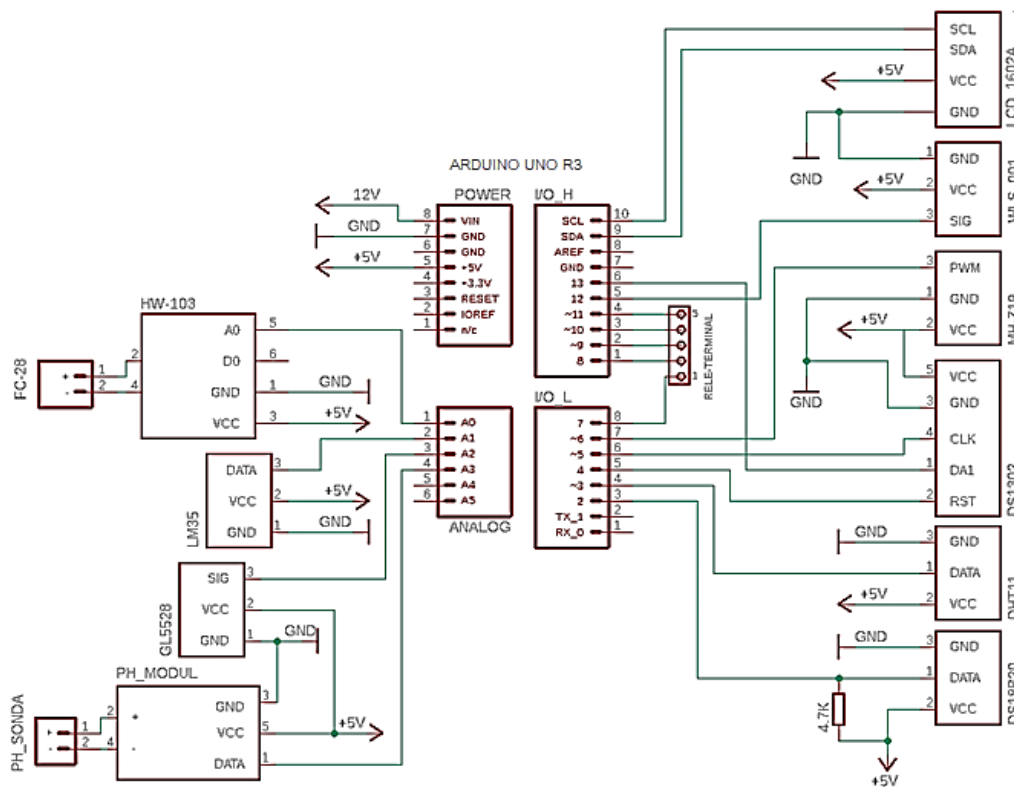
Pred samotným návrhom elektrickej schémy skleníka je potrebné definovať vstupy a výstupy riadiacej jednotky. Na vstupy riadiacej jednotky Arduina prichádzajú údaje zo sensorov. Na základe informácií spracovaných zo sensorov budú na výstupy odosielané príkazy na vypísanie údajov (pre LCD displej) a zapnutie, resp. vypnutie výstupných zariadení. Návrh blokovej schémy inteligentného skleníka s definovanými vstupmi a výstupmi je uvedený na Obr. 3.



Obr. 3 Návrh blokovej schémy inteligentného skleníka.

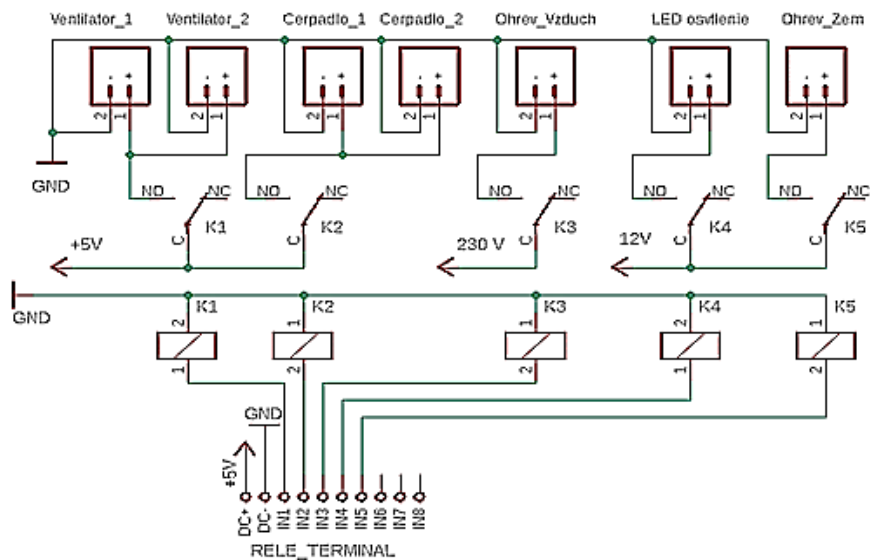
B. Elektrická schéma inteligentného skleníka

Inteligentný skleník bude napájaný sieťovým napätím 230 V. Arduino UNO R3 je napájané zo zdroja 230 V AC/12 V DC. Mimo Arduina sa 12 V DC napája aj topný kábel na ohrev zeminu a LED pásik. Pre ventilátory, čerpadlá a 8 kanálové relé s jumperom je potrebné napájanie 5 V (DC-DC menič 12/5V). Ventilátory, čerpadlá, LED pásik, ohrev zeminu a vzduchu sa budú zapínať/vypínať pomocou relé, na ktoré príde signál z Arduina. Schéma zapojenia vstupov a výstupov riadiacej jednotky Arduina UNO R3 je zobrazená na Obr. 4.



Obr. 4 Zapojenie vstupov a výstupov riadiacej jednotky Arduina UNO R3.

Schéma zapojenia koncových zariadení spínaných pomocou reléového bloku je zobrazená na Obr. 5, pričom spínanie je riadené z Arduina.



Obr. 5 Zapojenie koncových zariadení spínaných pomocou relé.

IV. REALIZÁCIA INTELIGENTNÉHO SKLENÍKA

A. Konštrukcia skleníka

Na konštrukciu skleníka sú použité drevotrieskové dosky, polykarbonátové dosky, hliníkové profily, lepidlá, skrutky, izolácie, pánty a podpery potrebné pre udržanie police, na ktorú sa umiestni samotný kvetináč. Doska na ktorej je elektronika je na koľajničkách, aby sa dala vysúvať. Montáž konštrukcie skleníka pri pohľade spredu a je zobrazená na Obr. 6. pri pohľade zozadu na Obr. 7.



Obr. 6 Montáž konštrukcie skleníka- pohľad spredu.



Obr. 7 Montáž konštrukcie skleníka- pohľad zozadu.

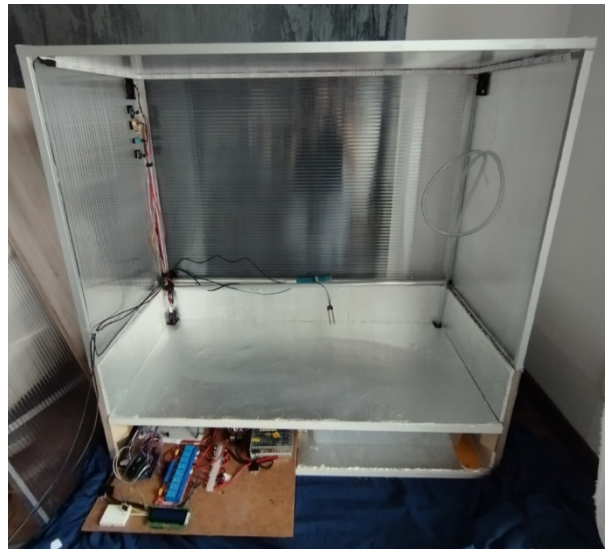
Spodná (drevená) časť skleníka spolu s policou je odizolovaná pomocou izolačnej fólie. Na ľavú stranu dna sú pripevnené koľajničky spolu s doskou, na ktorej bude osadená elektronika. Doska je vyberateľná pre ľahký prístup k elektronike. Do pravej a ľavej hornej časti polykarbonátových stien sú vyrezané otvory pre ventilátory. Do police, na ktorej bude kvetináč je zo zadnej pravej aj ľavej strany urobený štvorcový výrez, cez ktorý sú privedené do hornej časti skleníka senzory a hadice pre zavlažovanie. Finálna konštrukcia pred osadzovaním súčiastok je zobrazená na Obr. 8.



Obr. 8 Konštrukcia skleníka pred osadzovaním súčiastok.

B. Osadzovanie súčiastok

Rozloženie jednotlivých komponentov je navrhnuté tak, aby sa krížilo čo najmenej káblov. Zo zadnej steny je privedené napájacie napätie 230 V cez poistku do napájacieho zdroja. Z výstupov zdroja je ± 12 V cez svorkovnice napájané Arduino UNO R3, relé pre LED pásik, relé pre vykurovanie zeminy a DC menič napätia. Z výstupu (5 V) meniča napätia sa cez svorkovnicu napája relé pre ventilátory, čerpadlá a aj samotný vstup relé. Vykurovací kábel, ktorý slúži na ohrev vzduchu je priamo pripojený na svorky zdroja. Z periférií Arduina sa napájajú senzory, hodinový modul a displej. Senzory, ktoré je potrebné vyviesť, sú prepojené pomocou kontaktnej dosky s Arduino. Z kontaktnej dosky pokračujú rozvody káblov cez svorkovnicu na miesta inštalácie snímačov. Finálna realizácia zapojenia elektrickej časti skleníka je zobrazená na Obr. 9.



Obr. 9 Finálna realizácia zapojenia elektrickej časti skleníka.

Senzor teploty vzduchu, senzor CO₂, senzor vlhkosti vzduchu a senzor osvetlenia je pripojený do ľavej hornej časti skleníka pod ventilátor. Senzor hladiny vody spolu s čerpadlami je umiestnený v pravej časti skleníka v plastovej nádobe (Obr.8). Do vrchnáka plastovej nádoby sú vyrezané otvory pre snímač hladiny vody, prívod ku čerpadlám a na hadice, ktoré sú vyvedené do kvetináča.

Topný kábel slúžiaci na vykurovanie vzduchu je vložený do dutej hliníkovej rúrky a zasypaný je pieskom, ktorý po ohriatí bude udržiavať teplo. Topný kábel na vykurovanie zeme je tiež zasypaný do piesku na dno nádoby, do ktorej je potom nasypaná zemina pre posadenie rastlín. V nádobe je vložený senzor vlhkosti pôdy, senzor teploty pôdy a senzor na meranie pH pôdy. Hadice vedúce z čerpadiel sú uložené do 2 radov pozdĺž celej nádoby. Zo spodnej strany hadíc sú urobené otvory slúžiace na zavlažovanie.

Na záver pomocou kovových 90° závesov je pripojená predná drevotriesková stena a polykarbonátová stena pomocou kovových pántov. Pohľad na otvorený inteligentný skleník je zobrazený na Obr. 10.



Obr. 10 Inteligentný skleník.

C. Riadenie inteligentného skleníka.

Riadiaci program pre automatizovanú reguláciu v skleníku je písaný v programe open-source Arduino Software (IDE).

Celý riadiaci program sa skladá zo samostatných častí pre každý senzor, teda na vyčítanie

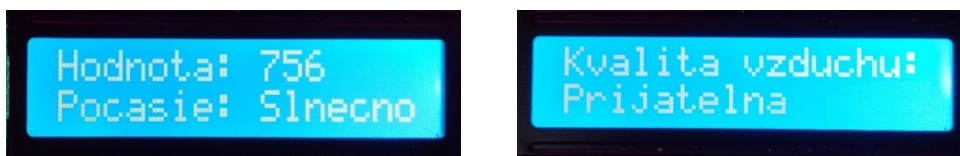
hodnôt z jednotlivých senzorov a z hodinového modulu. Pre každý senzor a modul sú zadané podmienky, čo sa ma stať na výstupe, ak na vstup prídu konkrétne údaje. Program obsahuje aj časť pre LCD displej na vypísanie informácií z jednotlivých senzorov, resp. modulov. Menu, ktoré sa zobrazuje na displeji je automatické, postupne každých 5 sekúnd sa menia na obrazovke hodnoty prichádzajúce z jednotlivých senzorov. Na nasledujúcich obrázkoch sú uvedené zobrazenia jednotlivých hodnôt senzorov.



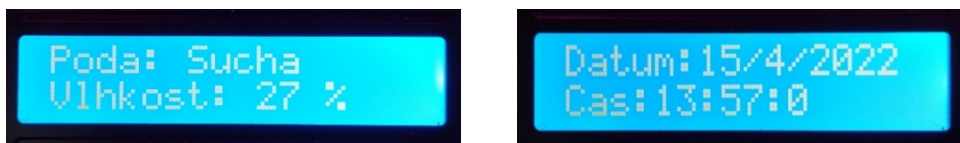
Obr. 11 Zobrazenie načítanej hodnoty z PH modulu a zo senzora teploty vzduchu.



Obr. 12 Zobrazenie načítanej hodnoty zo senzora vlhkosti vzduchu a zo senzora teploty pôdy.



Obr. 13 Zobrazenie načítanej hodnoty z fotorezistora a z CO2 senzora kvality ovzdušia.



Obr. 14 Zobrazenie načítanej hodnoty zo senzora vlhkosti pôdy a z hodinového modulu - dátum a čas.



Obr. 15 Zobrazenie načítanej hodnoty zo senzora hladiny vody.

V. ZÁVER

V tomto článku je prezentovaný konštrukčný návrh, realizácia a riadenie automatizovaného inteligentného skleníka. Pre načítavanie informácií zo snímačov a riadenie je použitá riadiaca jednotka Arduino UNO R3. Riadiaci program pre automatizovanú reguláciu v skleníku je písaný v programe open - source Arduino Software (IDE), ktorý podstatne uľahčuje písanie riadiaceho programu a jeho nahrávanie na dosku Arduina.

Po realizácii skleníka a jeho otestovaní riadenia môžeme skonštatovať, že výsledná konštrukcia spĺňa požiadavky inteligentného skleníka, no poskytuje aj priestor pre prípadné budúce vylepšenia. Je možnosť doinštalovania napríklad tienenia skleníka, ktorému sa doteraz nevenovala pozornosť, vzhľadom na to, že si to podmienky nevyžadovali. Taktiež je možnosť doinštalovania aj iných senzorických zariadení, ktoré by mohli zlepšiť kvalitu automatizovaného skleníka.

POĎAKOVANIE

Táto práca bola podporená Agentúrou na podporu výskumu a vývoja na základe zmluvy č. APVV-19-0210.

REFERENCES

- [1] A. Poncin, *What is a smart greenhouse?* [online]. 05. 12. 2021 [cit. 01. 03 .2022]. Dostupné na internete: < <https://ruuvi.com/what-is-a-smart-greenhouse-a-beginners-guide/>>.
- [2] *GGS. 4 Benefits of automated control systems in a commercial greenhouse.* [online]. 22.05.2022 [cit. 01. 03 .2022]. Dostupné na internete: < <https://ggs-greenhouse.com/blog/4-benefits-of-automated-control-systems-in-a-commercial-greenhouse>>.
- [3] *Smart greenhous monitoring systems.* [online]. 2021 [cit. 08. 03 .2022]. Dostupné na internete: < <https://images.squarespace-cdn.com/content/v1/5bfb50a87106999abcd2f7f2/1613601112407-OERSKIZJFSFNLPIU8ED3U/smart-sensors-solutions-banner.jpg?format=2500w>>.
- [4] Ben, *What is an arduino.* [online]. 09.11.2011 [cit. 08. 03 .2022]. Dostupné na internete: < <https://learn.sparkfun.com/tutorials/what-is-an-arduino/all>.
- [5] *Arduino UNO R3.* [online]. 13.08.2021 [cit. 08. 03 .2022]. Dostupné na internete: < <https://docs.arduino.cc/hardware/uno-rev3>>.
- [6] *LM35 Temperature sensor.* [online]. 27.08.2017 [cit. 08. 03 .2022]. Dostupné na internete: < <https://components101.com/sensors/lm35-temperature-sensor>>.
- [7] *DHT11 Temperature and Humidity Sensor Module.* [online]. 16.07.2021 [cit. 08. 03 .2022]. Dostupné na internete: < <https://components101.com/sensors/dht11-temperature-sensor>>.
- [8] S. Shah, *Soil moisture sensor arduino project.* [online]. 27.07.2021 [cit. 08. 03 .2022]. Dostupné na internete: < <https://arduinopoint.com/soil-moisture-sensor-arduino-project/>>.
- [9] *Interfacing DS18B20 1-wire Digital Temperature Sensor with Arduino.* [online]. 01.03.2019 [cit. 06. 04 .2022]. Dostupné na internete: < <https://lastminuteengineers.com/ds18b20-arduino-tutorial/>>.
- [10] *Chewett. Light Intensity Sensor Module 5528 Photo Resistor Arduino review and code.* [online]. 05.04.2022 [cit. 08. 03 .2022]. Dostupné na internete: < <https://chewett.co.uk/blog/389/light-intensity-sensor-module-5528-photo-resistor-arduino-review-code/>>.
- [11] *Arduino CO2 Sensor – MH-Z19 Veispiel und Sketch.* [online]. 17.09.2021 [cit. 08. 03 .2022]. Dostupné na internete: < <https://iot-space.dev/arduino-co2-sensor-mh-z19-beispiel-und-sketch/>>.
- [12] *Safepools pH.* [online]. 01.09.2016 [cit. 08. 03 .2022]. Dostupné na internete: < <https://create.arduino.cc/projecthub/safepools-ph/safepools-ph-408029>>.
- [13] *Arduino lesson water sensor.* [online]. 27.09.2017 [cit. 08. 03 .2022]. Dostupné na internete: < <https://osoyoo.com/2017/09/27/arduino-lesson-water-sensor/>>.
- [14] *Topný kábel 12V, 5M.* [online]. 23.01.2020 [cit. 06. 04 .2022]. Dostupné na internete: < <https://www.fonetip.sk/topny-kabel-12-v-5-m/dC6224472/>>.
- [15] *Step down buck converter interface adjustable power supply module board.* [online]. 13.06.2020 [cit. 18. 04 .2022]. Dostupné na internete: < <http://www.icstation.com/step-down-buck-converter-interface-adjustable-power-supply-module-board-24v12v-p-12228.html/>>.
- [16] *Hodinový modul DS1302.* [online]. 14.09.2019 [cit. 06. 04 .2022]. Dostupné na internete: < <https://techfun.sk/produkt/hodinovy-modul-ds1302/>>.
- [17] *Modry LCD display 16x2 I2C.* [online]. 02.01.2021 [cit. 18. 04 .2022]. Dostupné na internete: < <http://https://www.gme.sk/modry-lcd-display-16x2-i2c>>.
- [18] *Zdroj TLP-12-75 (trafo) 12V 150W - vnútorný.* [online]. 15.09.2020 [cit. 06. 04 .2022]. Dostupné na internete: < <https://www.123led.sk/led-zdroj-trafo-12v-150w-vnutorny/>>.
- [19] *Ventilátor 2-pinový.* [online]. 09.02.2021 [cit. 06. 04 .2022]. Dostupné na internete: < <https://techfun.sk/produkt/ventilator-2-pinovy-5v/>>.
- [20] *Malá vodná pumpa.* [online]. 11.07.2020 [cit. 06. 04 .2022]. Dostupné na internete: < <https://techfun.sk/produkt/mala-vodna-pumpa/>>.
- [21] *8 kanálové relé s jumperom H/L.* [online]. 28.10.2021 [cit. 06. 04 .2022]. Dostupné na internete: < https://techfun.sk/produkt/8-kanalove-rele-s-jumperom-hl/?attribute_pa_rele=5-v/>.
- [22] *LED grow pasik plnospektralne 12W-m.* [online]. 15.09.2020 [cit. 06. 04 .2022]. Dostupné na internete: < <https://www.ledmegrow.sk/led-grow-pasik-plnospektralne-12w--m/>>.
- [23] *DFRobot SEN0161 Gravity.* [online]. 15.01.2021 [cit. 08. 03 .2022]. Dostupné na internete: < <https://www.rapidonline.com/dfrobot-sen0161-gravity-analog-ph-sensor-meter-kit-for-arduino-75-0249/>>.

Nepriamy jednosmerný menič s jednoduchým odľahčovacím obvodom na sekundárnej strane

¹Marek PÁSTOR, ²Jaroslav DUDRIK, ³Richard MICHAL

^{1,2,3}Katedra elektrotechniky a mechatroniky, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Košice, Slovenská republika

¹marek.pastor@tuke.sk, ²jaroslav.dudrik@tuke.sk, ³richard.michal@student.tuke.sk

Abstrakt — Článok sa zaoberá analýzou, simuláciou a meraním na vysokofrekvenčnom jednosmernom meniči s mäkkým spínaním a rekuperačnou odľahčovacou sieťou sa sekundárnej strane. Menič má mostíkovú topológiu a je riadený modifikovaným šírkovým riadením. Na sekundárnej strane obsahuje riadený uzlový usmerňovač s jedným tranzistorom. Opísaný spôsob riadenia odstraňuje okruhový prúd meniča a zabezpečuje zapínanie v nule napätia a vypínanie v nule prúdu pre primárne tranzistory v plnom rozsahu záťaže. Navrhnutý odľahčovací obvod zabezpečuje vypínanie v nule napätia pre sekundárny tranzistor a rekuperáciu energie do záťaže. Činnosť meniča je overená meraním na laboratórnom modeli s maximálnym výkonom 2 kW a spínacou frekvenciou 100 kHz.

Kľúčové slová — jednosmerný menič, mäkké spínanie, odľahčovací obvod

I. ÚVOD

Účinnosť a rozmery meniča sú dôležitými parametrami pri návrhu jednosmerného meniča. Zmenšovanie rozmerov je možné dosiahnuť zvýšením spínacej frekvencie meniča, čo má však za následok nárast spínacích strát meniča a s tým spojený pokles účinnosti meniča [1]. J jednosmerný menič v mostíkovom zapojení riadený šírkovým riadením a fázovým posunom je v praxi využívaný kvôli pevnej spínacej frekvencii a jednoduchej realizácii mäkkého spínania primárnych tranzistorov formou zapnutia v nule napätia. To sa dosahuje využitím rozptylovej indukčnosti transformátora a výstupných kapacít tranzistorov v širokom rozsahu zaťaženia meniča. Avšak tento menič nedokáže dosahovať zapínanie v nule napätia pre celý rozsah záťaže. Pri odľahčení meniča už nie je energia nahromadená v rozptylovej indukčnosti dostatočná na prebitie výstupných kapacít tranzistorov v zaostávajúcej vetve meniča. Ďalším nedostatkom spomínaného meniča je prítomnosť okruhového prúdu tečúceho počas úseku nulového napätia na primárnej strane. To má za následok zvýšenie vodivostných strát meniča.

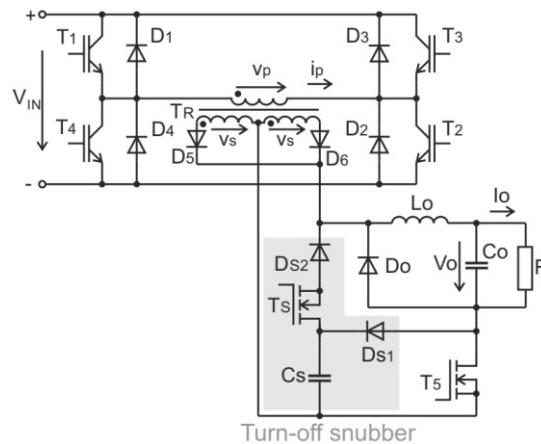
Pre odstránenie okruhového prúdu a rozšírenie rozsahu záťaže, v ktorom jednosmerný menič v mostíkovom zapojení dosahuje zapínanie v nule napätia boli publikované rôzne zapojenia [5] – [8]. Vyznačujú sa použitím viac alebo menej komplexného prídavného obvodu na primárnej alebo sekundárnej strane meniča. Prínos týchto zapojení však často limitovaný hodnotou prúdu záťaže a pri malých záťažiach sú často neúčinné. Iným spôsob odstránenia spomínaných problémov predstavuje použitie riadeného usmerňovača. Významnú časť okruhového prúdu tvorí prúd sekundárnej strany meniča pretransformovaný na primárnu stranu. Riadený usmerňovač umožňuje odpojenie sekundárnej strany a tým aj prerušenie okruhového prúdu. V kombinácii s vhodným návrhom meniča je možné dosiahnuť zapínanie primárnych tranzistorov v nule napätia nezávisle od veľkosti prúdu záťaže len s využitím magnetizačného prúdu transformátora. Odľahčovací obvod zabezpečí mäkké spínanie tranzistora riadeného usmerňovača a pri vhodnom návrhu je zároveň možné dosiahnuť rekuperáciu energie odľahčovacieho obvodu do záťaže.

II. PRINCÍP ČINNOSTI MENIČA

A. Topológia meniča

Opisovaný jednosmerný menič pozostáva z mostíkového striedača (tranzistory $T_1 - T_4$ a antiparalelné diódy $D_1 - D_4$), vysokofrekvenčného znižovacieho transformátora TR, riadeného

usmerňovača (diódy D_5 a D_6 a tranzistor T_5), výstupného filtra (induktor L_O a kapacitor C_O) a nulovej diódy D_0 (Obr. 1). Sekundárny tranzistor T_5 umožňuje prerušiť okruhový prúd a zároveň zabezpečuje, že primárne tranzistory vypínajú len magnetizačný prúd transformátora. Spínacie straty tranzistora usmerňovača T_5 sú minimalizované navrhnutým odľahčovacím obvodom s rekuperáciou energie. Ten pozostáva z tranzistora T_S , dvoch diód D_{S1} a D_{S2} a kapacitora C_S .



Obr. 1 Schéma zapojenia meniča

B. Spôsob riadenia jednosmerného meniča

Tranzistory striedača sú radené širkovým riadením s frekvenciou 100 kHz a konštantným fázovým posunom 180° . Tranzistor usmerňovača je spínaný dvojnásobnou frekvenciou, 200 kHz a premenlivou striedou. To umožňuje riadiť tok výkonu v meniči. Tranzistor usmerňovača je vypínaný pred vypnutím tranzistorov striedača. Prúd na primárnej strane transformátora je tak znížený na hodnotu magnetizačného prúdu a primárne tranzistory vypínajú v nule prúdu. Tranzistor odľahčovacej T_S siete má rovnakú spínaciu frekvenciu ako tranzistor usmerňovača T_5 . Činnosť navrhnutého jednosmerného meniča môže byť rozdelená do 10 intervalov zobrazených na Obr. 2.

Interval (t_0 - t_1): Primárne tranzistory T_1 a T_2 sú spolu s tranzistorom usmerňovača T_5 zapnuté v čase t_0 . Tým sa spustí komutácia prúdu záťaže z nulovej diódy D_0 na usmerňovaciu diódu D_5 . Strmosť nárastu prúdu na primárnej a sekundárnej strane je limitovaná rozptylovou indukčnosťou transformátora. To zabezpečuje zapínanie tranzistorov T_1 , T_2 a T_5 v nule prúdu.

Interval (t_1 - t_2): Komutácia medzi nulovou diódou D_0 a usmerňovacou diódou D_5 sa ukončí v čase t_1 . Prúd tranzistora T_5 dosiahol hodnotu prúdu vyhladzovacej tlmivky L_O .

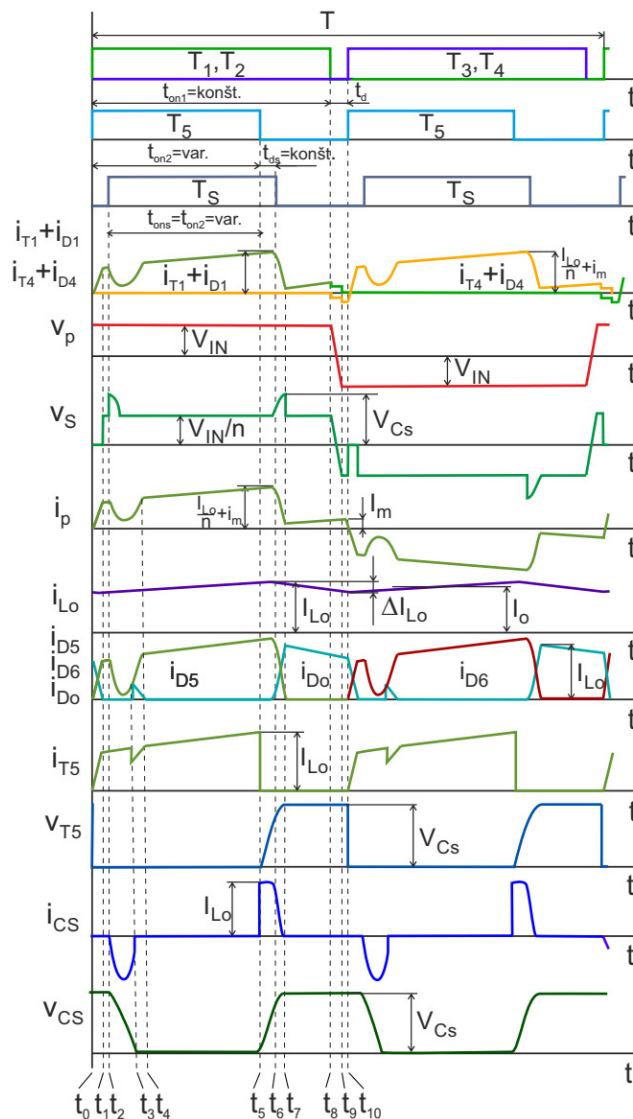
Interval (t_2 - t_3): V čase t_2 sa zapne odľahčovací tranzistor T_S . To spustí komutáciu medzi diódou D_5 a tranzistorom T_S . Táto komutácia prebieha rezonančným kmitom. Strmosť nárastu prúdu tranzistora T_S je tak limitovaná a zapína v nule prúdu. Zapnutie tranzistora T_S pripojí odľahčovací kapacitor C_S na výstup usmerňovača, čo spôsobí odľahčenie primárnych tranzistorov.

Interval (t_3 - t_4): V čase t_3 dôjde k vybitiu odľahčovacieho kapacitora C_S . Prúd záťaže tak komutuje späť na nulovú diódu D_0 . Odľahčovací tranzistor T_S je vypnutý v čase t_3 v nule prúdu.

Interval (t_4 - t_5): Na primárnej strane vedú prúd tranzistory T_1 a T_2 a na sekundárnej dióda D_5 . Toto je hlavný interval prenosu energie z primárnej strany na sekundárnu.

Interval (t_5 - t_6): V čase t_5 vypne sekundárny tranzistor T_5 . Prúd záťaže komutuje cez odľahčovaciu diódu D_{S1} na odľahčovací kapacitor C_S . Nárast kolektorového napätia tranzistora T_5 je limitovaný strmosťou nabíjania odľahčovacieho kapacitora C_S a tranzistor T_5 tak vypína v nule napätia. Maximálne kolektorové napätia tranzistora T_5 je limitované napätím kapacitora C_S .

Interval (t_6 - t_7): V čase t_7 dosiahne napätie kapacitora C_S hodnotu usmerného napätia. To polarizuje diódu D_0 v priepustnom smere a spustí sa komutácia prúdu z usmerňovacej diódy D_5



Obr. 2 Časové intervaly jednosmerného meniča s mäkkým spínaním

na nulovú diódu D_0 . Obvod komutácie obsahuje rozptylovú indukčnosť transformátora a tá sa jej energia odovzdá do odľahčovacieho kapacitора C_s . To spôsobí nárast sekundárneho napätia nad hodnotu usmerneného napätia definovanú napájacím napätím a prevodom transformátora.

Interval (t_7 - t_8): Primárny prúd v čase t_7 poklesol na hodnotu magnetizačného prúdu transformátora. Nedochádza k prenosu energie z primárnej strany na sekundárnu stranu meniča. Celý prúd záťaže je dodávaný výstupným LC filtrom.

Interval (t_8 - t_9): V čase t_8 vypnú primárne tranzistory. Keďže viedli len malý magnetizačný prúd, tak ich vypnutie je v nule prúdu. Výstupné kapacity primárnych tranzistorov T_1 a T_2 sú nabíjané magnetizačným prúdom transformátora. To obmedzuje strmú nárastu ich kolektorového napätia, a tak tranzistory vypínajú zároveň v nule napätia. Celý proces prebývania výstupných kapacít sa musí udiť v rámci bezpečnostnej medzery t_d .

Interval (t_9 - t_{10}): Magnetizačný prúd zároveň vybíja výstupné kapacity tranzistorov T_3 a T_4 . Pri ich úplnom vybití je v nasledovnej pol perióde zabezpečené zapnutie tranzistorov T_3 a T_4 v nule napätia.

C. Rekuperačný odľahčovací obvod

Riadený usmerňovač umožňuje odstrániť dva najväčšie nedostatky meničov so širkovým riadením s fázovým posunom. Okruhový prúd je úplne odstránený vypnutím usmerňovača skôr, než dôjde k vypnutiu tranzistorov na primárnej strane. Tým sa výrazne znižujú vodivostné straty striedača. Druhou výhodou použitia riadeného usmerňovača je nezávislosť dosahovania zapínania v nule napätia od prúdu záťaže pre obe vetvy striedača. Na prebývanie výstupných

kapacít tranzistorov striedača sa využíva len magnetizačný prúd transformátora. Avšak pre dosiahnutie mäkkého spínania tranzistora riadeného usmerňovača musí byť zapojenie doplnené o odľahčovací obvod na sekundárnej strane. Odľahčovací obvod musí odstrániť spínacie straty tranzistora usmerňovača, pohltiť energiu rozptylovej indukčnosti transformátora a zabezpečiť vracanie pohltenej energie späť do záťaže. Zložitosť odľahčovacieho obvodu by mala byť čo najnižšia. Odľahčovací obvod prezentovaný v tomto článku pozostáva len z jedného tranzistora T_S , dvoch diód D_{S1} a D_{S2} a jedného kapacitora C_S . Odľahčovací obvod neobsahuje induktor, nakoľko využíva rozptylovú indukčnosť transformátora. Zároveň zabezpečuje odovzdávanie energie do záťaže meniča, čím je ho možné považovať za bezstratový.

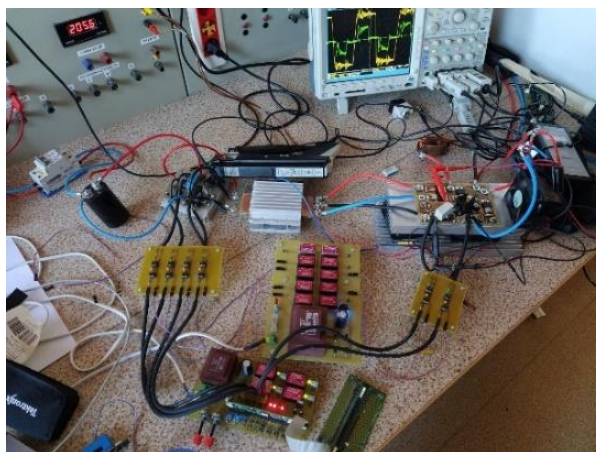
III. LABORATÓRNY MODEL MENIČA

Pre overenie činnosti navrhnutého meniča bol vyrobený laboratórny model s výkonom 2 kW zobrazený na Obr. 3. Základné parametre meniča sú uvedené v tabuľke 1. Parametre vysokofrekvenčného transformátora sú uvedené v tabuľke 2 a komponenty meniča sú uvedené v tabuľke 3. Pre účely predbežného overenia činnosti meniča bol laboratórny model napájaný napätím 200 V a prúd záťaže bol nastavený na 30 A. Spínacia frekvencia striedača bola 100 kHz a spínacia frekvencia tranzistora usmerňovača bola nastavená na 200 kHz.

Činnosť primárneho tranzistora T_1 je znázornená na Obr. 4. Tranzistor T_1 zapína v nule napätia a zároveň je magnetizačný prúd transformátora vedený diódou D_1 . Tranzistor T_1 tak zapína v nule napätia a v nule prúdu, teda má ideálne podmienky pre zapínanie. Z Obr. 3 je taktiež zrejмый pokles kolektorového prúdu tranzistora T_1 po zapnutí tranzistora T_S , čím dochádza k zníženiu efektívnej hodnoty prúdu tranzistora T_1 a zníženiu jeho vodivostných strát. Sekundárny tranzistor T_S je vypnutý pred vypnutím tranzistora T_1 , čo má za následok pokles primárneho prúdu na hodnotu magnetizačného prúdu transformátora, ktorý je následne vypnutý tranzistorom T_1 . Strmosť nárastu kolektorového napätia tranzistora T_1 je limitovaná nabíjaním jeho výstupnej kapacity magnetizačným prúdom transformátora počas bezpečnostnej medzery. Tranzistor T_1 tak vypína v nule prúdu a v nule napätia, čím dosahuje ideálne podmienky pre vypínanie IGBT tranzistorov.

Tabuľka 1
Základné parametre meniča

Parameter	Hodnota
Napájacie napätie U	300 V
Výstupný prúd I_o	50 A
Výstupné napätie naprázdno U_o	50 V
Spínacia frekvencia striedača f_s	100 kHz
Bezpečnostná medzera t_d	1 μ s
Odľahčovací kapacitor C_S	150 nF



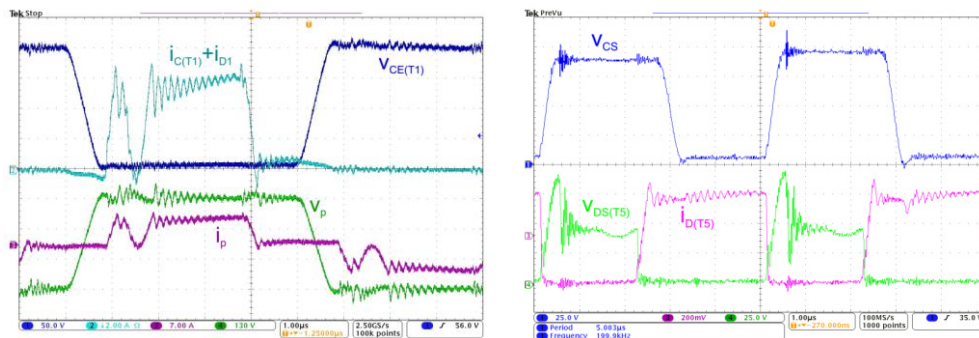
Obr. 3 Laboratórny model meniča

Tabuľka 2
Parametre vysokofrekvenčného transformátora

Parameter	Hodnota
Prevod p	1:5
Hlavná indukčnosť L_m	650 μH
Rozptylová indukčnosť na primárnej strane L_σ	3 μH
Pracovná frekvencia f_s	100 kHz
Amplitúda magnetizačného prúdu I_m	1,2 A

Tabuľka 3
Komponenty laboratórneho modelu meniča

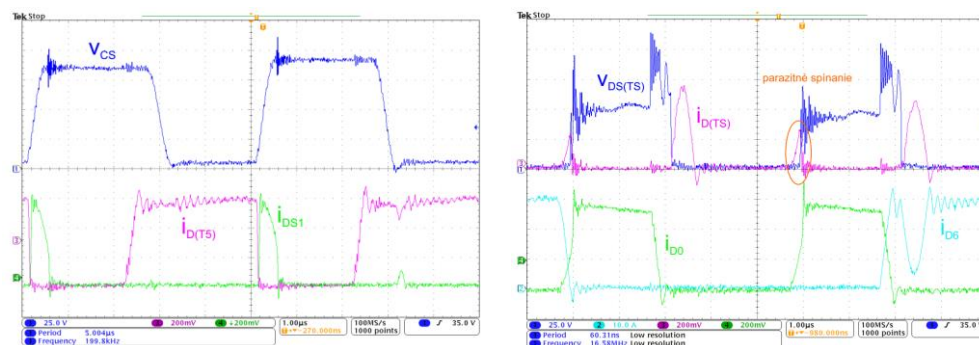
Parameter	Hodnota
Primárne tranzistory $T_1 - T_4$	G4PSC71UD
Usmerňovacie diódy D_5 a D_6 a diódy odľahčovacieho obvodu D_{S1} D_{S2}	150EBU04
Tranzistor usmerňovača T_5 a odľahčovacieho obvodu T_S	IRFP4668
Výstupný filter L_O/C_O	5 $\mu\text{H}/150\mu\text{F}$
Odľahčovací kapacitor C_S	100 nF



a) kolektorové napätie tranzistora T_1 $v_{CE(T1)}$ (50 V/d.) a kolektorový prúd tranzistora T_1 a diódy D_1 $i_{C(T1)+i_{D1}}$ (2 A/d.) (hore) a primárne napätie v_p (130 V/d.) a primárny prúd i_p (7 A/d.) transformátora (dole).

b) napätie odľahčovacieho kapacitora v_{CS} (25 V/d.) (hore) a kolektorové napätie tranzistora T_5 $v_{DS(T5)}$ (25 V/d.) a kolektorový prúd $i_{D(T5)}$ (10 A/d.) sekundárneho tranzistora T_5 (dole).

Obr. 4 Časové priebehy meniča



a) napätie odľahčovacieho kapacitora v_{CS} (25 V/d.) (hore) a kolektorový prúd sekundárneho tranzistora $i_{D(T5)}$ (10 A/d.) a prúd odľahčovacej diódy D_{S1} i_{DS1} (10 A/d.) (dole).

b) kolektorové napätie odľahčovacieho kapacitora $v_{DS(T5)}$ (25 V/d.) a jeho kolektorový prúd $i_{D(T5)}$ (10 A/d.) (hore) a prúd nulovej diódy i_{D0} (10 A/d.) a usmerňovacej diódy D_5 i_{DS} (10 A/d.) (dole).

Obr. 5 Časové priebehy sa sekundárnej strane meniča

Vypnutie sekundárneho tranzistora pred vypnutím primárnych tranzistorov síce zabezpečí odstránenie spínacích strát na primárnej strane meniča, ale presunie ich na sekundárnu stranu. Pre ich úplné odstránenie je navrhnutý odľahčovací obvod, ktorý zabezpečí presun spínacích strát do záťaže meniča. Činnosť odľahčovacieho obvodu je znázornená na Obr. 5. Pri vypnutí tranzistora T_5 skokom preberie jeho prúd odľahčovací kapacitor C_S cez diódu D_{S1} . Prúd záťaže spôsobí nabitie kapacitora C_S na hodnotu výstupného napätia. Kapacitor C_S zároveň pohlcuje energiu rozptylovej indukčnosti transformátora, čo sa prejaví nárastom napätia na kapacitore C_S (Obr. 4 b)). Strmosť nárastu kolektorového prúdu tranzistora T_5 je limitovaná rozptylovou indukčnosťou transformátora. Tranzistor T_5 tak zapína v nule prúdu a vypína v nule napätia. Spínanie odľahčovacieho tranzistora T_5 je ovplyvnené parazitnými osciláciami (Obr. 5 b)). Aj napriek tomu však tranzistor T_5 zapína v nule prúdu v dôsledku rezonančného kmitu medzi odľahčovacím kapacitorom C_S a rozptylovou indukčnosťou transformátora. Počas tohto kmitu sú oba sekundárne tranzistory otvorené, čo zabezpečí cestu pre odovzdanie energie uloženej v kapacitore C_S do záťaže meniča.

IV. ZÁVER

V článku je opísaný nepriamy jednosmerný menič s mäkkým spínaním a odľahčovacím obvodom na sekundárnej strane. Princíp činnosti meniča je overený meraním na laboratórnom modeli meniča pri znížených parametroch. Meraním je dokázaná základná činnosť meniča. Primárne tranzistory meniča zapínajú a vypínajú v nule prúdu. Tým je zabezpečené ich mäkké spínanie. Taktiež je v meniči odstránený okruhový prúd, ktorý je typický pre meniče so širkovým riadením s fázovým posuvom. Navrhnutý odľahčovací obvod na sekundárnej strane zabezpečuje vypínanie tranzistora usmerňovača v nule napätia, pohltenie energie rozptylovej indukčnosti transformátora a jej odovzdanie do záťaže. Ďalším krokom je optimalizácia kapacity odľahčovacieho kapacitora a odstránenie parazitných oscilácií pri spínaní tranzistora T_5 .

REFERENCIE

- [1] P. Spanik, B. Dobrucky, M. Frivaldsky, P. Drgona, I. Kurytnik, "Measurements of Switching Losses in a Power Transistor Structure", *Elektronika ir Elektrotechnika*, Vol. 2, 2008, pp. 75–78
- [2] H. Aksoy and M. T. Aydemir, "Comparison of zero voltage switching phase-shifted PWM full-bridge DC-DC converter topologies," Intl Aegean Conference on Electrical Machines & Power Electronics (ACEMP), 2015 Intl Conference on Optimization of Electrical & Electronic Equipment (OPTIM) & 2015 Intl Symposium on Advanced Electromechanical Motion Systems (ELECTROMOTION), 2015, pp. 818-824
- [3] G. N. B. Yadav and N. L. Narasamma, "An Active Soft Switched Phase-Shifted Full-Bridge DC-DC Converter: Analysis, Modeling, Design, and Implementation," in *IEEE Transactions on Power Electronics*, vol. 29, no. 9, pp. 4538-4550, Sept. 2014
- [4] Lee, Il-Oun & Moon, Gun-Woo, "Phase-Shifted PWM Converter With a Wide ZVS Range and Reduced Circulating Current", *IEEE Transactions on Power Electronics*, Vol. 28. 2013, pp. 908-919.
- [5] J. Dudrik, M. Pastor, M. Lacko, R. Zatkovic, "High frequency Sift-Switching PWM DC-DC Converter With Active Output Rectifier Operating as a Current Source for Arc Welding Applications," *Electric Power Components and Systems*, Vol. 45, No. 6, 2017, pp. 681–691
- [6] J. Dudrik, M. Pastor, M. Lacko, R. Zatkovic, "Zero-Voltage nad Zero-Current Switching PWM DC-DC Converter Using Secondary Rectifier With one Active Switch and Nondissipative Turn-Off Snubber", *IEEE Transactions on Power Electronics*, Vol. 33, No. 7, 2018, pp. 6012–6023
- [7] J. Dudrik, M. Bodor, V. Ruscin, "Zero-voltage and zero-current switching DC-DC converter with controlled output rectifier," in *Proc. 14th International Power Electronics and Motion Control Conference*, Ohrid, 2010, pp. T1-39–T1-44.
- [8] M. Leso, J. Zilkova, P. Girovsky, "Development of Simple Fuzzy Logic Controller for DC-DC Converter", *Proc. IEEE 18th International Power Electronics and Motion Control Conference (PEMC)*, 2018, pp. 86–93

Necieľové kontradiktórne útoky na klasifikačné hlboké modely

¹Dmytro LAHUNOV, ²Ivan ČÍK, ³Marián MACH

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

¹dmytro.lahunov@student.tuke.sk, ²ivan.cik@tuke.sk, ³marian.mach@tuke.sk

Abstrakt — Článok sa zameriava na problematiku kontradiktórnych útokov na hlboké modely, keď pomerne malou zmenou spracovávaných vstupných dát je možné dosiahnuť zmenu rozhodnutia hlbokého modelu oproti rozhodnutiu na nemodifikovaných dátach. Pozornosť je zameraná na konvolučné modely pre klasifikáciu obrazov. Článok sa zaoberá necieľovými kontradiktórными útokmi, pričom sa hlbšie venuje dvom metódam útoku – metóde FGSM (Fast Gradient Sign Method) a metóde BIM (Basic Iterative Method). V rámci experimentovania bolo realizovaných niekoľko experimentov s týmito metódami v rámci dvoch dátových množín – jednoduchšej MNIST doméne a zložitejšej CIFAR-10 doméne.

KLúčové slová — hlboký model, kontradiktórna perturbácia, kontradiktórny útok, Fast Gradient Sign Method, Basic Iterative Method

I. ÚVOD

V oblasti používania modelov vytvorených strojovým učením je veľmi dôležitým faktorom dôvera vo výsledky, poskytované týmito modelmi na základe vstupných dát (napr. obrazov v prípade modelov založených na hlbokom učení pre klasifikáciu objektov).

Cieľom kontradiktórnych útokov je snaha zmiasť naučený model a prinútiť ho produkovať nevalidné rozhodnutia [1]. Sú založené na modifikácii vstupných dát – pridávaní malých perturbácií k vstupným dátam, ktoré tieto dáta zmenia určitým spôsobom. Príkladom sú zmeny jasových úrovní jednotlivých pixelov, z ktorých je vytváraný vstupný obraz. Tieto modifikácie sú pre ľudské oko buď nerozpoznateľné alebo automaticky kompenzované bez toho aby si to človek uvedomoval. Pre človeka tak modifikovaný obraz zobrazuje tú istú informáciu ako pôvodný nemodifikovaný obraz. Pre naučený model však diferenciacia medzi modifikovaným a pôvodným obrazom môže byť významná a dokonca môže viesť k rôznym rozhodnutiam. Pre človeka je potom zmena rozhodnutia modelu pri modifikovanom obraze nepochopiteľná. Dôsledkom môže byť strata dôvery ľudí v daný model alebo činnosť automatických systémov odlišná od ich predpokladaného chovania.

Algoritmy kontradiktórnych útokov sa delia na dve hlavné kategórie – cieľové a necieľové kontradiktórne útoky [2]. Útoky, ktoré majú snahu „pokaziť“ model aby produkoval odlišné výsledky pre kontradiktórne príklady voči výsledkom pre pôvodné príklady sa označujú ako necieľové. Na druhej strane cieľové útoky sa snažia ovplyvniť model nielen na získanie odlišných výsledkov pre kontradiktórne príklady ale aj na určenie aké by tieto odlišné výsledky mali byť.

Tento článok sa zameriava na oblasť necieľových kontradiktórnych útokov v oblasti klasifikácie obrazov. Jeho štruktúra je nasledovná – v druhej kapitole sa venuje vybranej dvojici metód vytvárania necieľových útokov a v tretej kapitole je uvedená séria experimentov s danými metódami na dvoch dátových množinách.

II. NECIEĽOVÉ KONTRADIKTÓRNE ÚTOKY

Hlboký model sa typicky učí takým spôsobom, že učiaci proces sa snaží minimalizovať chybovú funkciu, ktorá vyjadruje rozdiel medzi skutočným výsledkom, ktorý by mal byť dosiahnutý (napr. triedou do ktorej by mal byť vstupný obraz klasifikovaný), a výsledkom predikovaným daným modelom. Teda vstupný obraz X klasifikovaný do triedy Y_{true} je spracovávaný modelom a výsledkom je predikcia do triedy Y_{pred} .

$$f(Y_{pred}, Y_{true}) \quad (1)$$

Necieľový kontradiktórny útok mení tento obraz na X_{adv} takým spôsobom, aby chybová funkcia (1) bola nie minimalizovaná ale naopak maximalizovaná, teda aby sa pravdepodobnosť predikovaného výsledku čo najviac líšila od skutočného. Zmena X na X_{adv} sa označuje ako kontradiktórna perturbácia, ktorá môže byť jednokroková alebo viackroková.

A. Metóda FGSM (Fast Gradient Sign Method)

FGSM [3, 4] je algoritmom, ktorý predstavuje základ necieľových kontradiktórnych útokov. Je to jednoduchý algoritmus, pri ktorom sa perturbácie k originálnemu obrazu pridávajú iba v jednom kroku. Kontradiktórny obraz generovaný týmto algoritmom sa vytvára podľa nasledujúceho vzťahu

$$X_{adv} = X + \epsilon * \text{sign}(\nabla_X J(X, Y_{true})) \quad (2)$$

teda kontradiktórny príklad nie je zameraný na maximalizáciu pravdepodobnosti predikcií pre konkrétnu cieľovú triedu, perturbácie sa vypočítajú pomocou funkcie *sign* ktorá dostáva ako svoj parameter gradient chybovej funkcie, ktorý sa dá vypočítať napríklad pomocou algoritmu späťného šírenia chyby.

Na konci sa hodnota funkcie *sign* ešte vynásobí tzv. koeficientom perturbácií ϵ . Úroveň pridaných perturbácií je regulovaná hodnotou tohto koeficienta.

Táto metóda je veľmi intuitívna, je navrhnutá na využitie prvku, ktorý je použitý pre učenie siete - gradientu. Útok využíva gradient chybovej funkcie s ohľadom na vstupné dáta (pôvodný obraz) a pomocou neho upravuje vstupné dáta pre maximalizáciu chybovej funkcie.

B. Metóda BIM (Basic Iterative Method)

BIM [5] je taktiež reprezentantom necieľového kontradiktórneho útoku. Je viackrokovým algoritmom, namiesto pridávania perturbácie v jednom kroku tento algoritmus pridáva perturbácie vo viacerých krokoch. BIM je priamym rozšírením algoritmu FGSM – pridáva špeciálny krokový parameter α , ktorý určuje hodnotu perturbácie pridávanej v jednom iteračnom kroku. BIM funguje podľa nasledujúceho rekurzívneho vzťahu

$$X_{adv}^0 = X, X_{adv}^{N+1} = \text{Clip}_{X, \epsilon} \{X_{adv}^N + \alpha * \text{sign}(\nabla_X J(X_{adv}^N, Y_{true}))\} \quad (3)$$

kde *Clip* označuje ohraničenie hodnoty na dodržanie maximálnej perturbácie pre každý pixel obrazu. Je definovaná ako

$$\min(255, X + \epsilon, \max(0, X - \epsilon, X')) \quad (4)$$

a teda udržiava výslednú hodnotu vo vnútri intervalu $[X-\epsilon, X+\epsilon]$. Hodnotu parametra α je doporučované nastaviť na 1/255 (na jednotku intenzity pixelu) a počet iterácií tak, aby pixel mohol dosiahnuť maximálnu hodnotu perturbácie ϵ .

Činnosť tohto algoritmu je podobná ako pri (2) ibaže sa vykonáva opakovane počas niekoľkých iterácií. Namiesto jedného veľkého kroku vykonáva sériu krokov veľkosti α . Tým pádom je metóda BIM pomalšia ako metóda FGSM, avšak vo všeobecnosti je úspešnejšia v produkcii kontradiktórnych obrazov.

III. VYBRANÉ EXPERIMENTY S KONTRADIKTÓRNymi ÚTOKMI

A. Testovací problém

Pre experimentovanie s rôznymi metódami kontradiktórnych útokov boli zvolené dve dátové množiny obrazov:

- MNIST – množina šedotónových obrazov rozmeru 28x28 pixelov. Obrazy reprezentujú ručne písané číslice 0 až 9 (každý obraz obsahuje iba jednu z číslic), pričom sú použité rôzne formy písma a sklon číslic.
- CIFAR-10 – farebné obrazy kódované pomocou RGB farebného modelu, obrazy sú rozmeru 32x32 pixelov. Obrazy reprezentujú desať rôznych tried: koň, pes, mačka, auto, nákladné auto, loď, lietadlo, vták, žaba, jeleň.

V oboch prípadoch sa jednalo o klasifikáciu obrazov do jednej z možných klasifikačných tried, použitých v danej dátovej množine.

B. Hlboké modely pre klasifikáciu obrazov

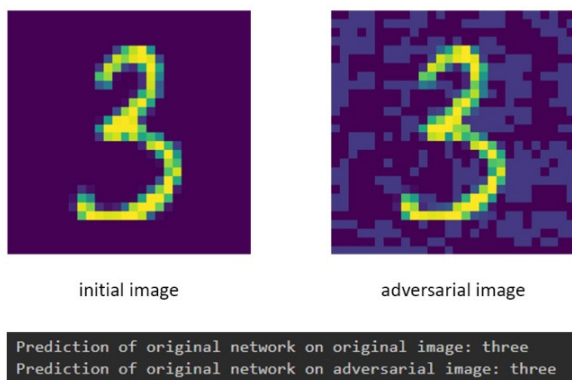
Pre účely klasifikácie obrazov použitých dátových množín boli použité konvolučné siete. Pretože obtiažnosť klasifikácie je rôzna pre použitie uvažovaných dátových množín, pre každú

z nich bola použitá iná štruktúra konvolučnej siete (jednoduchšia pre MNIST a zložitejšia pre CIFAR-10). Tieto štruktúry boli:

- Pre MNIST – tri konvolučné vrstvy (s 32, 64 a 64 výstupnými kanálmi), poolingová vrstva (max, jadro 2x2), tri plne prepojené vrstvy (s 32, 32 a 10 výstupnými neurónmi) doplnené dropout vrstvami pre zníženie nebezpečenstva preučenia.
- Pre CIFAR-10 – šesť konvolučných vrstiev (s 32, 32, 64, 64, 128 a 128 výstupnými kanálmi), dve poolingové vrstvy (obe max s jadrom 2x2), tri plne prepojené vrstvy (so 128, 128 a 10 výstupnými neurónmi) doplnené dropout vrstvami a vrstvami pre dávkovú normalizáciu.

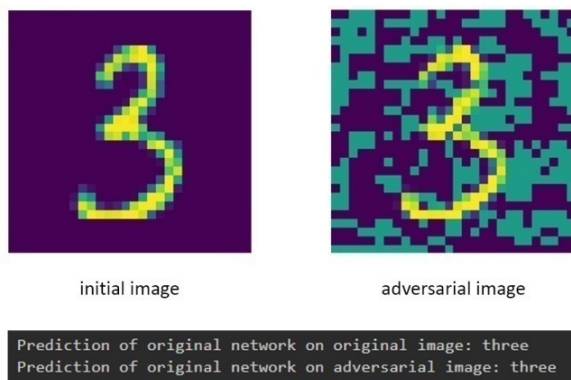
C. Experiment s algoritmom FGSM

Metóda má iba jeden parameter - koeficient perturbácií ϵ , na ktorom závisí intenzita pridávaných perturbácií. Teda podstata experimentov je v skúmaní vplyvu veľkosti tohto parametra na výsledok klasifikácie modelu [6]. Začneme s databázou MNIST. Pre $\epsilon=0.1$ dostaneme



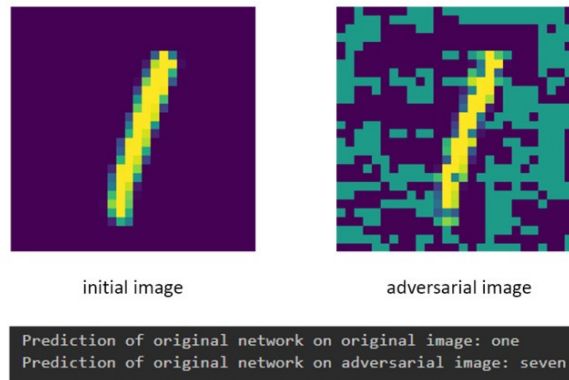
Obr. 1 MNIST, $\epsilon=0.1$

Na obr. 1 vidno pôvodný a kontradiktórny obrázok spolu s ich klasifikáciou naučeným modelom. Vidno, že model je do určitej miery robustný, pre danú hodnotu perturbácií sa nedal zmiatť. Zvýšime hodnotu na $\epsilon=0.6$ a skúsime zaútočiť znova.



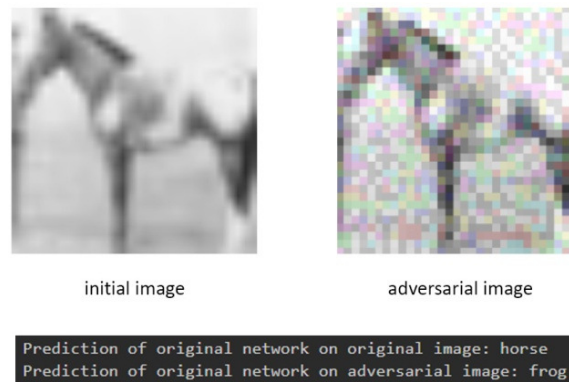
Obr. 2 MNIST, $\epsilon=0.6$

Napriek zvýšenej miere perturbácií (viditeľne zvýšenej pri porovnaní kontradiktórnych obrazov na obr. 1 a obr. 2) sa predikcia nezmenila. Avšak ak pri tej istej hodnote koeficienta sa použije iná číslica, tak predikcia je zmenená – obr. 3.



Obr. 3 MNIST, $\epsilon=0.6$

Očividne, pre každý obraz existuje iná hodnota pre daný koeficient ϵ , ktorá už spôsobuje zmenu klasifikácie kontradiktórnej podoby obrazu voči originálnej. Pre overenie metódy použijeme aj druhú dátovú množinu a dostaneme

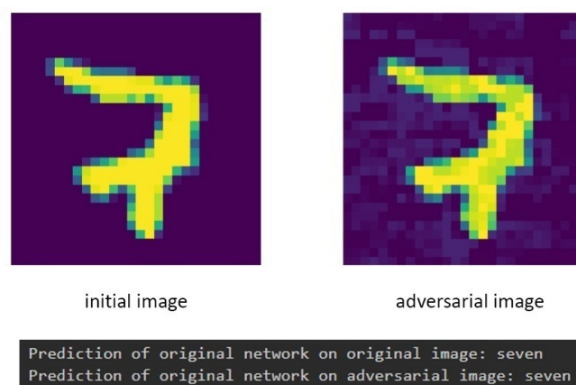


Obr. 4 CIFAR-10, $\epsilon=0.06$

Keďže obrázky v druhej dátovej množine sú komplexnejšie než v prvej, na realizáciu kontradiktórnej zmeny, dostatočne veľkej pre zmenu klasifikácie, postačuje menšia hodnota koeficientu ϵ .

D. Experiment s algoritmom BIM

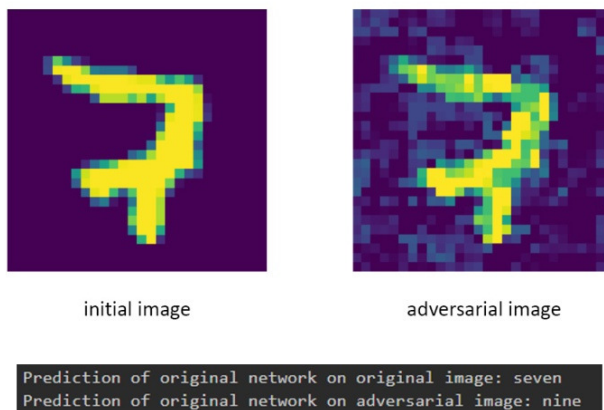
Pre účely testovania [6] tento algoritmus vyžaduje nastaviť hodnoty troch rozličných parametrov: krokový parameter α , maximálnu perturbáciu ϵ a počet iterácií. Začneme s kombináciou $\epsilon=0.1$, počtom iterácií 20 a $\alpha=0.1/20$. Výsledkom je



Obr. 5 MNIST, $\epsilon=0.1$ #=20, $\alpha=0.1/20$

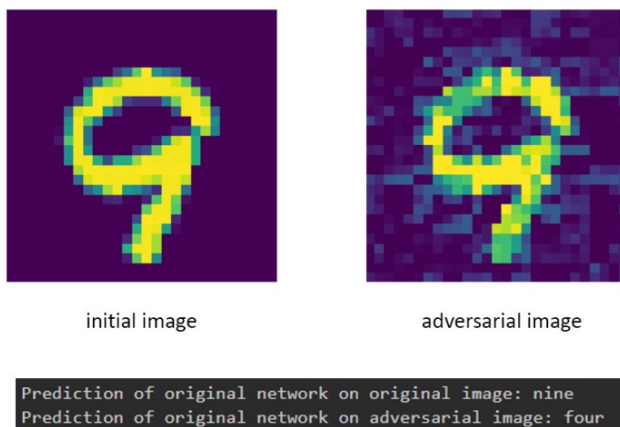
Podľa obr. 5 vidno, že malá hodnota ϵ nepôsobí na zmenu klasifikácie ani v prípade, že perturbácie sú pridávané opakovane počas niekoľkých iterácií. Jej zvýšenie (spolu so zmenou hodnôt ostatných parametrov s cieľom zachovať malú hodnotu zmeny počas jednej iterácie)

vedie na



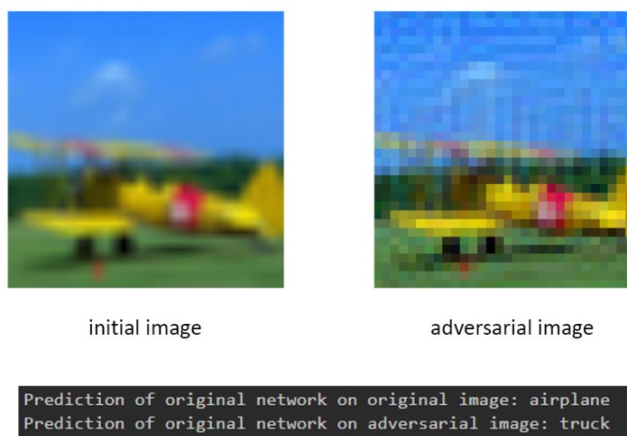
Obr. 6 MNIST, $\epsilon=0.3$ $\#=40$, $\alpha=0.3/40$

Viditeľnosť perturbácií sa o niečo zvýšila (ale nie až tak výrazne ako to bolo nutné pre metódu FGSM na obr. 2 a obr. 3) a výstup klasifikačného modelu sa zmenil. Rovnaké hodnoty parametrov použijeme na inú číslicu a dostaneme



Obr. 7 MNIST, $\epsilon=0.3$ $\#=40$, $\alpha=0.3/40$

Kombinácia hodnôt parametrov nebola špecializovaná na pokazenie klasifikácie jedného typu číslice ale je účinná aj pre iný typ. Záverečne ešte metódu overíme na databáze CIFAR-10, pričom opäť použijeme menšiu hodnotu maximálnej perturbácie: $\epsilon=0.05$, počet iterácií 40 a $\alpha=0.05/40$. Výsledkom je opäť zmenená klasifikácia kontradiktórneho obrazu voči klasifikácii obrazu pôvodného.



Obr. 8 CIFAR-10, $\epsilon=0.05$ $\#=40$, $\alpha=0.05/40$

IV. ZÁVER

Vykonané experimenty potvrdili, že klasifikačné hlboké modely založené na štandardných konvolučných sieťach je možné zmiast' a malou zmenou klasifikovaných obrazov, ktorá by nerobila problém ľuďom, je možné dosiahnuť nevalidnú klasifikáciu týmito modelmi. Pri jednoduchých obrazoch je síce nutné použiť kontradiktórnu perturbáciu väčšieho rozsahu, avšak v prípade komplexnejších obrazov stačí aj menší rozsah pridaných perturbácií na dosiahnutie želaného efektu.

Je zrejmé, že citlivosť hlbokých modelov na kontradiktórne útoky môže byť v praxi problémom a preto je potrebné sa venovať problematike zvyšovania robustnosti týchto modelov a tým pádom aj ich ochrane pred kontradiktórnymi útokmi.

LITERATÚRA

- [1] J. Shaw, "AI and Adversarial Attacks", *Harvard Magazine*, January-February 2019.
- [2] A. Chakraborty et al., "A survey on adversarial attacks and defences", *CAAI Trans. on Intelligence Technology*, roč. 6, 2021, 25-45.
- [3] I. J. Goodfellow, J. Shlens a C. Szegedy, "Explaining and Harnessing Adversarial Examples", Int. Conference on Learning Representations ICLR 2015. Dostupné: <https://arxiv.org/pdf/1412.6572.pdf>
- [4] M. Andriushchenko a N. Flammarion, "Understanding and Improving Fast Adversarial Training", 34th Conference on Neural Information processing NeurIPS, 2020, Vancouver, Canada.
- [5] A. Kurakin, I. J. Goodfellow a S. Bengio, "Adversarial examples in the physical world", Int. Conference on Learning Representations ICLR 2017. Dostupné: <https://arxiv.org/pdf/1607.02533.pdf>
- [6] D. Lahunov, *Štúdium algoritmov vysvetliteľnej umelej inteligencie: kontradiktórne útoky*, Bakalárska práca, Fakulta elektrotechniky a informatiky Technická univerzita v Košiciach, 2021.

Prvé experimenty s hracím agentom založeným na algoritme MCTS

¹Dávid LACKO, ²Marián MACH

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

¹david.lacko.4@student.tuke.sk, ²marian.mach@tuke.sk

Abstrakt — Článok sa zameriava na problematiku vytvárania herných agentov pre hranie antagonistických hier dvoch hráčov s nulovým súčtom. Pozornosť je zameraná na algoritmus MCTS (Monte Carlo Tree Search) vhodný pre realizáciu efektívneho kontradiktórneho prehľadávania herného stromu v spojitosti s realizáciou náhodných simulácií priebehu hry. V článku sú uvedené výsledky experimentov s hracím agentom na báze MCTS pričom pozornosť je venovaná ako vybraným parametrom algoritmu tak aj porovnaniu výkonu hracieho agenta voči trom rôznym protihráčom s rôznym stupňom hracích schopností. Všetky experimenty boli realizované pre hru Tic-Tac-Toe.

KLúčové slová — antagonistická hra, herný strom, kontradiktórne prehľadávanie, MCTS, simulácia hry, Tic-Tac-Toe

I. ÚVOD

Dôležitú kategóriu hier tvoria antagonistické hry hráčov s perfektnou informáciou. Teda hry, ktorých stav je dostupný každému zúčastnenému hráčovi – každý hráč má všetky informácie o priebehu a aktuálnom stave hry. Tieto hry môžu byť deterministické (napr. šach, dáma) ale tak isto môžu obsahovať aj nedeterministické chovanie (napr. backgammon). Majú konkurenčný charakter, zisk jedného hráča je kompenzovaný stratou iného hráča. Takéto hry sú hry s nulovým súčtom – súčet ziskov/strát všetkých hráčov je nulový.

Možné priebehy hry sú reprezentované vetvami herného stromu. Uzly tohto stromu reprezentujú jednotlivé stavy hry, pričom koreň stromu je tvorený počiatočným stavom hry. Hrany stromu reprezentujú možné ťahy hráčov – ak z jedného uzla vedie hrana do druhého uzla, potom ak v stave, reprezentovanom prvým uzlom, hráč vykoná ťah reprezentovaný danou hranou, tak hra prechádza do nového stavu, ktorý je reprezentovaný druhým uzlom. Jeden stav môže byť reprezentovaný viacerými uzlami (ak sa do tohto stavu dá dostať viacerými sekvenciami vykonaných ťahov). Čím dlhšia sekvencia ťahov vedie k nejakému stavu, tým je uzol, reprezentujúci tento stav po vykonaní danej sekvencie ťahov hlbšie vnorený v strome. Listové uzly herného stromu reprezentujú stavy, v ktorých hra končí a je v nich možné určiť výsledok hry.

Jedna cesta v hernom strome od koreňového uzla k listovému uzlu reprezentuje jednu možnú realizáciu hry. Preto tento strom je možné použiť ako model hry, umožňujúci určovať ťahy hráča prehľadávaním stromu. Použitá metóda prehľadávania však musí realizovať kontradiktórne prehľadávanie, zohľadňujúce striedanie jednotlivých hráčov – hráči majú totiž protichodné ciele.

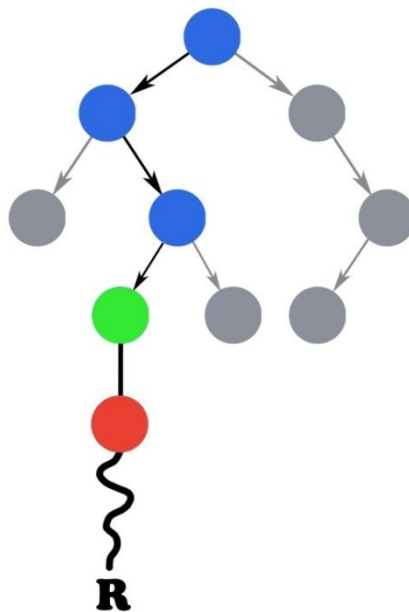
Klasickou metódou kontradiktórneho prehľadávania je minimax [1], ktorý sa v hernej teórii používa na nájdenie optimálneho ťahu hráča. Metóda stojí na predpoklade, že protihráč hrá tiež optimálne. Jeden hráč (označovaný ako Max) sa snaží získať čo najvyššie skóre, zatiaľ čo protihráč (označovaný ako Min) sa snaží dosiahnuť pravý opak. Minimax algoritmus vlastne vykonáva prehľadávanie herného stromu slepým prehľadávaním. Nevýhodou sú veľké výpočtové nároky – samotný algoritmus je veľmi jednoduchý, avšak úplné prehľadávanie veľkých herných stromov nie je prakticky realizovateľné. Pre zníženie nárokov je možné použiť alfa-beta orezanie [2], ktoré umožňuje neprehľadávať niektoré časti herného stromu bez vplyvu na výsledok prehľadávania. Účinnosť orezavania je však nepredikovateľná, v najhoršom možnom prípade k žiadnemu orezaniu ani nedôjde.

II. ALGORITMUS MCTS

Algoritmus MCTS (Monte Carlo Tree Search) [3] kombinuje prehľadávanie herného stromu so simulovaním priebehu hry. Pre toto využíva simuláciu Monte Carlo na zhromažďovanie odhadov hodnôt, ktoré vedú k vysoko výnosným trajektóriám v tomto strome. Keďže nepracuje s úplným herným stromom ale iba jeho časťou (typicky iba malým fragmentom stromu, ktorý inkrementálne zväčšuje), je výpočtovo efektívny a preto použiteľný aj pre rozsiahlejšie herné stromy.

Základná štruktúra algoritmu MCTS je pomerne jednoduchá, herný strom je budovaný inkrementálne a to asymetrickým spôsobom. Každá jedna iterácia algoritmu pozostáva zo štyroch krokov:

- výber,
- rozšírenie,
- simulácia,
- spätné šírenie.



Obr. 1 Ilustrácia inkrementálneho budovania herného stromu algoritmom MCTS.

Ilustrujme postup budovania herného stromu pomocou obr. 1, pričom predpokladajme, že na začiatku iterácie algoritmu je herný strom tvorený iba modrými, zeleným a šedými uzlami (teda bez červeného uzla), pričom každý z uzlov má nejaké aktuálne ohodnotenie.

Výber je vlastne zostup herným stromom od koreňového uzla do jedného z uzlov s doposiaľ nepreskúmaným ťahom, pričom sa použije politika výberu potomka v prípadoch, keď k zostupu z nejakého uzla existuje viacero alternatív. Na obrázku je zostup značený cestou prechádzajúcou modrými uzlami a končiacou v zelenom uzle, ktorý je síce aktuálne listovým uzlom budovaného stromu avšak nie je listovým uzlom celého herného stromu, pretože v stave, reprezentovanom daným uzlom, hra môže ešte pokračovať.

Rozšírenie realizuje pridanie nového uzla. V stave, ktorý je reprezentovaný zeleným uzlom, sa zistia všetky dostupné ťahy, jeden z nich sa vyberie a po jeho vykonaní sa hra dostane do nového stavu, ktorý bude reprezentovaný pridaným červeným uzlom.

Simulácia slúži na ohodnotenie novo pridaného uzla. Ak tento uzol ešte nie je koncovým stavom hry, nie je ho možné ohodnotiť podľa toho, ako hra skončila. Preto sa realizuje simulácia ďalšieho pokračovania hry až do koncového stavu, keď je možné ohodnotiť dosiahnutý výsledok R . Pri samotnej simulácii sú jednotlivé ťahy volené náhodne, simuláciu je možné vykonať jednu alebo niekoľko.

Spätné šírenie výsledku R umožňuje aktualizovať ohodnotenie uzlov na ceste od koreňového až po červený uzol, pričom sa to deje v opačnom poradí. Hodnota červeného uzla bude odvodená od R , potom bude aktualizovaná hodnota zeleného uzla, následne modrého predchodcu zeleného uzla, atď.

A. Politika výberu potomka

V minulosti často používanou politikou je UCB (Upper Confidence Bounds) [4]. Každý uzol už známeho fragmentu má priradenú UCB hodnotu a pri výbere potomka sa vyberá ten uzol, ktorý má túto hodnotu najvyššiu. Táto hodnota je definovaná ako

$$UCB_i = x_i + \sqrt{\frac{2 \log N}{n_i}} \quad (1)$$

kde x_i je stredná hodnota ohodnotenia uzla i , n_i je počet simulácií pre uzol i , a N je počet simulácií pre všetky rodičovské uzly.

Možno vidieť vo vzťahu (1), že čím je uzol viackrát navštívený, tým sa druhá časť stáva menšou a pravdepodobnosť výberu uzla klesá. Dôsledkom je, že šanca sa dáva aj iným uzlom. Na druhej strane, čím je hodnota uzla vyššia, tým väčší dôraz sa kladie na jeho výber.

V [5] bolo UCB kritérium rozšírené pre kontradiktórne prehľadávanie herných stromov pod menom UCT (UCB applied to Trees) definované ako

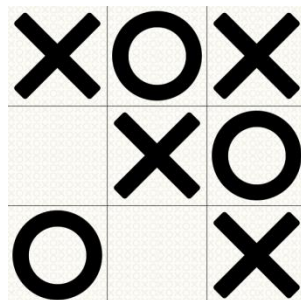
$$UCT_i = x_i + 2c \sqrt{\frac{2 \log N}{n_i}} \quad (2)$$

kde n_i je počet, koľkokrát bol navštívený uzol i , N je počet, koľkokrát bol navštívený nadradený uzol k uzlu i a c je nezáporná konštanta. V prípade $n_i = 0$ je UCT miera považovaná za nekonečne veľkú, čo má za následok že uzly, ktoré ešte neboli navštívené, sú vybrané prednostne.

III. VYBRANÉ EXPERIMENTY S MCTS HRACÍM AGENTOM

A. Testovací problém a metodika testovania

Pre experimentovanie bola zvolená známa hra Tic-Tac-Toe, pri ktorej sa ťahovo striedajú dvaja hráči pri umiestňovaní svojich značiek na dosku s rozmerom 3x3 polia. Vyhráva ten z hráčov, ktorému sa ako prvému podarí umiestniť tri svoje značky horizontálne, vertikálne alebo diagonálne. Ukážka hry je na obr. 2.



Obr. 2 Hracia doska Tic-Tac-Toe s ukážkou umiestnenia značiek oboch hráčov.

Táto hra je deterministická s plne pozorovateľným prostredím, pričom môže končiť výhrou jedného z hráčov alebo remízou, keď ani jeden z hráčov nie je schopný realizovať výherný ťah. Pre túto hru neexistuje výherná stratégia, ktorá by garantovala niektorému z hráčov výhru pri jej striktnom nasledovaní.

Bolo vytvorené hracie prostredie [6], ktoré umožňovalo pripojiť dvoch hracích agentov. Toto prostredie umožňuje jednak realizáciu jednej hry medzi agentmi a taktiež aj realizáciu turnaja medzi nimi. Pri turnaji obaja agenti hrajú určitý počet hier, pričom sa pravidelne striedajú v úlohe začínajúceho hráča. Prostredie vyhodnocuje ich výsledky a poskytuje údaje o úspešnosti jednotlivých agentov v zmysle početnosti výhier niektorého agenta alebo remíz.

Všetky experimenty boli realizované v tomto prostredí ako turnaje pozostávajúce z 1000 hier dvoch agentov. V niektorých experimentoch hrali proti sebe dve verzie MCTS agenta (s rozličnými alebo rovnakými charakteristikami), v iných zase hral MCTS agent voči automatickému botovi, simulujúcemu hráča s vopred definovanými schopnosťami využívať niektoré vybrané taktické obraty danej hry.

B. Porovnanie parametrov

Cieľom prvého experimentu bolo porovnať dve politiky výberu potomka použité v prvej fáze iterácie algoritmu pri výbere cesty od koreňového uzla (staršiu UCB a novšiu UCT) a zistiť, ktorá z nich je vhodnejšia. Dosažené výsledky sú v tab. 1.

Tabuľka 1
UCB výber vs UCT výber.

Počet cyklov učenia	Výhra UCT [%]	Výhra UCB [%]	Remíza [%]
100	19.4	20.3	60.3
1000	27.2	1.1	71.7

Pri malom počte učiacich cyklov sú obe politiky porovnateľné, prekvapivo s miernou prevahou UCB. Avšak situácia je diametrálne odlišná v prípade dlhšej doby učenia, keď dochádza k výraznejšej degradácii výkonu UCB. Očividne dochádza k preučeniu, ktoré je v prípade UCB zreteľne významnejšie než v prípade UCT.

Cieľom druhého experimentu bolo porovnať dva spôsoby rozširovania vybraného uzla herného stromu o podradený uzol v druhej fáze iteračného cyklu. Pri náhodnom rozšírení sa podradený uzol vyberie náhodne, každý podradený uzol má rovnakú pravdepodobnosť. Pri deterministickom rozširovaní sa uzly vyberajú v dopredu danom poradí. Dosiahnuté výsledky sú v tab. 2.

Tabuľka 2
Náhodné rozšírenie vs deterministické rozšírenie.

Počet cyklov učenia	Náhodné [%]	Deterministické [%]	Remíza [%]
100	18.4	19.4	62.2
1000	26.2	1.3	72.5

Opäť pri malom počte učiacich cyklov sú obe metódy porovnateľné, avšak pri väčšom počte očividne vhodná voľba poradia skúmania jednotlivých dostupných ťahov hrá podstatnú úlohu. A keďže vhodné poradie nie je známe, je lepšie preferovať náhodný výber pred výberom deterministickým.

C. Hra proti sebe samému

Cieľom experimentu bol výkon MCTS agenta v hre proti sebe samému – teda hra dvoch rovnocenných protivníkov. Bola použitá UCT politika s náhodným rozšírením. Dosiahnuté výsledky sú v tab. 3.

Tabuľka 3
MCTS agent vs MCTS agent.

Počet cyklov učenia	Výhra MCTS1 [%]	Výhra MCTS2 [%]	Remíza [%]
100	19.7	19.6	60.7
1000	1.9	2.0	96.1

Vidno, že skutočne pre skúmanú hru neexistuje garantujúca výherná stratégia a teda protihráč má vždy možnosť dosiahnuť remízu. Dáta súčasne potvrdzujú význam väčšieho počtu učiacich cyklov a zároveň signalizujú, že aj ďalšie zvýšenie tohto počtu by mohlo mať zmysel, pretože agent má stále ešte rezervu vo svojej výkonnosti.

D. Hra proti jednoduchému hráčovi

Cieľom je porovnanie hracieho agenta, založeného na MCTS, voči niekoľkým hracím botom, simulujúcim hráčov s rôznymi schopnosťami hrania zvolenej hry.

Prvý experiment bol zameraný na hru proti náhodnému protivníkovi, ktorý svoje akcie vyberá náhodným spôsobom pričom každá z aktuálne dostupných akcií má rovnakú pravdepodobnosť výberu. Dosiahnuté výsledky sú v tab. 4.

Tabuľka 4
MCTS agent vs náhodný bot.

Počet cyklov učenia	Výhra MCTS [%]	Výhra bot [%]	Remíza [%]
100	84.0	3.6	12.4
1000	88.4	0.5	11.1

Potvrdil sa predpoklad, že MCTS vysoko prevyšuje náhodný výber ťahov. So zvýšením

počtu učiacich cyklov rozdiel medzi nimi ešte narástol, keď pravdepodobnosť výhry náhodného hráča poklesla na skoro zanedbateľnú hranicu.

Druhý experiment bol zameraný na hru proti jednokrovému hráčovi. Tento hráč v prípade, že existuje akcia ktorej vykonanie znamená výhru hráča, deterministicky volí práve túto výhernú akciu. V prípade, že takáto akcia neexistuje, vyberá svoj ťah náhodným spôsobom spomedzi aktuálne dostupných ťahov. Dosiahnuté výsledky sú v tab. 5.

Tabuľka 5
MCTS agent vs jednokrovový bot.

Počet cyklov učenia	Výhra MCTS [%]	Výhra bot [%]	Remíza [%]
100	80.0	11.3	8.7
1000	88.4	1.7	9.9

Aj v prípade hry proti hráčovi, ktorý simuluje začiatočníka v hre oboznámeného s pravidlami a schopného realizovať jednoduchý taktický cieľ výhry jedným ťahom, sa ukázala prevaha MCTS agenta. Jednokrovový hráč síce dosiahol lepší výsledok v počte výhier než predchádzajúci náhodný hráč, avšak počet výhier dosiahnutý MCTS agentom v prípade poskytnutia dostatočne dlhej doby učenia nepoklesol.

Tretí experiment bol zameraný na hru proti dvojkrovému hráčovi. Tento hráč v prípade, že existuje akcia ktorej vykonanie znamená výhru hráča, tiež deterministicky volí túto akciu. V prípade, že takáto akcia pre neho neexistuje ale naopak existuje pre protihráča, tak deterministicky volí akciu (ak taká existuje), ktorá zabráni protihráčovi vykonať daný vyhrávajúci ťah. V ostatných prípadoch vyberá svoj ťah náhodným spôsobom spomedzi aktuálne dostupných ťahov. Dosiahnuté výsledky sú v tab. 6.

Tabuľka 6
MCTS agent vs dvojkrovový bot.

Počet cyklov učenia	Výhra MCTS [%]	Výhra bot [%]	Remíza [%]
100	24.7	14.1	61.2
1000	28.5	5.6	65.9

Keďže schopnosti hráča simulovaného botom opäť narástli (je schopný nielen realizovať výhru jedným ťahom ale aj predvídať túto výhru u protihráča a brániť jej – neprehrá ak sa nedostane do situácie, že protihráč má k dispozícii dva výherné ťahy súčasne), prirodzene stúpol počet dosiahnutých remíz. Aj keď schopnosť MCTS agenta vyhrávať poklesla, naďalej prevažuje výkonnosť simulovaného protihráča, pričom naďalej platí že so zvýšeným počtom učiacich cyklov sa jeho schopnosť vyhrávať zlepšuje.

IV. ZÁVER

Vykonané experimenty potvrdili, že algoritmus MCTS pre kontradiktórne prehľadávanie herného stromu je vhodným základom pre realizáciu hracieho agenta, určeného pre hranie antagonistických hier dvoch hráčov s nulovým súčtom a úplnou informáciou. Algoritmus dokáže ťažiť zo zvýšeného počtu učiacich cyklov, avšak v tomto prípade je nutnosťou použiť stratégiu UCT, staršia stratégia UCB sa v tomto prípade ukázala ako nevhodná.

Otázne je, do akej miery dokáže víťaziť aj nad hráčmi s vyššou hernou kvalifikáciou, pretože vykonané experimenty sa zameriavali iba na hráča schopného rozpoznať a reagovať na jednokrovové taktické obraty, neboli skúmané viackrovové taktické obraty ani strategický prístup hráča k plánovaniu ťahov. Pre toto sú potrebné ďalšie experimenty v tejto oblasti.

LITERATÚRA

- [1] S. Russell, P. Norvig, *Artificial Intelligence: A modern approach*, Fourth edition, Pearson, 2020.
- [2] H.R. Ekbia, *Artificial Dreams: A Quest for Non-Biological Intelligence*, Cambridge University Press, 2008.
- [3] G. Chaslot et al, "Progressive Strategies for Monte-Carlo Tree Search", *New Mathematics and Natural Computation*, roč. 4, č. 3, 2008, 343-357.
- [4] C. Browne et al, "A Survey of Monte Carlo Tree Search methods", *IEEE Trans. on Computational Intelligence and AI in Games*, roč. 4, č. 1, 2012, 1-43.
- [5] L. Kocsis, C. Szepesvári, "Bandit-based Monte-Carlo Planning", *Proc. of the 17th European conference on machine learning*, Berlin, Springer, 2006, 282-293.
- [6] D. Lacko, *Hrací agent založený na MCTS (Monte Carlo Tree Search)*, Bakalárska práca, Fakulta elektrotechniky a informatiky Technická univerzita v Košiciach, 2021.

Skúmanie malvéru a neškodného softvéru v dokumentoch PDF

Daniel MAJIRSKÝ, Anton BALÁŽ, Branislav MADOŠ

¹Katedra počítačov a informatiky, Fakulta elektrotechniky a informatiky,
Technická univerzita v Košiciach

daniel.majirsky@student.tuke.sk, anton.balaz@tuke.sk, branislav.mados@tuke.sk

Abstract – Článok sa zaoberá problematikou skúmania malvéru a neškodného softvéru v dokumentoch formátu PDF a je rozdelený do niekoľkých častí. Prvá časť sa zaoberá popisáním existujúcich prístupov k detekovaniu a analýze malvéru. Následne je v článku popísaný návrh a implementácia riešenia na skúmanie malvéru v dokumentoch PDF pomocou klasterizácie. Pre tú sú použité algoritmy Kmeans a SOM. Výsledný systém pomáha používateľovi určiť, či predmetný dokument obsahuje škodlivý softvér alebo je bezpečný.

Keywords – analýza malvéru, PDF, klasterovanie, VirusTotal, Kmeans, samoorganizujúce sa mapy

I. ÚVOD

Dôvera ľudí v internetové služby narastá a s tým aj pokusy o ich zneužívanie, najčastejšie s finančnou motiváciou. Počítačoví útočníci stále rozširujú a vyvíjajú spôsoby šírenia škodlivého softvéru a získavania citlivých dát. Šírenie škodlivých alebo poškodených dokumentov patrí medzi stále preferované metódy šírenia malvéru nič netušiacemu používateľovi. S touto metódou sa taktiež spája sociálne inžinierstvo, pomocou ktorého sú príjemcovia takýchto dokumentov motivovaní, alebo dokonca prinútení tieto dokumenty otvoriť. Najbežnejšími formátmi týchto dokumentov sú PDF a Office Open XML, ktoré boli vytvorené práve pre poskytnutie globálnej platformy nezávislej od hardvérovej konfigurácie. Pre ich popularitu medzi bežnými používateľmi a kvôli prístupu do štruktúry dokumentov, sú tieto formáty veľmi atraktívne pre kyberútočníkov.

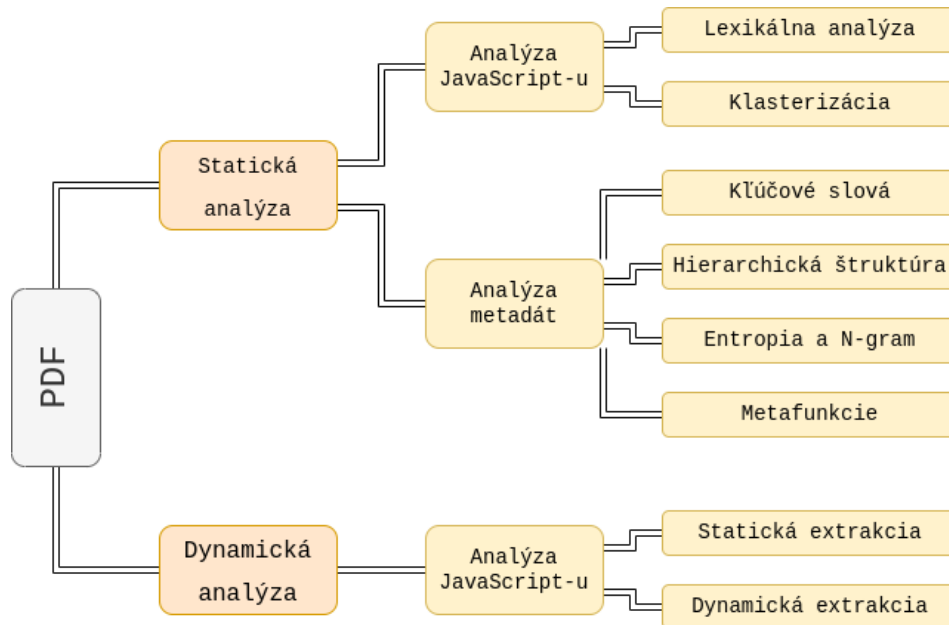
Malvérové infekcie boli v posledných rokoch jedným z hlavných problémov bezpečnostnej komunity. S rastúcim počtom nových škodlivých programov sa výrobcovia antivírusových softvérov snažia držať krok, s cieľom chrániť aj rastúci počet používateľov počítačov. Detekcia je stále viac a viac komplexná kvôli zvýšenej zložitosti škodlivého kódu, a tak je potrebné preskúmať pokročilejšie techniky na detekciu a boj proti týmto hrozbám.

Hlavným cieľom tejto práce je navrhnutie a implementácia systému na spracovanie a analýzu malvéru v dokumentoch formátu PDF. Riešenie tejto práce by malo byť schopné spracovať súbor vzoriek, na ktorých vykoná analýzu malvéru. Na základe získaných dát z analýzy by mal navrhovaný systém uplatniť metódu klasterovania dát, pomocou ktorej by malo byť možné zoskupiť dokumenty s podobnými znakmi obsahu. Od navrhovaného systému sa vyžaduje, aby predstavoval jednoduché konzolové riešenie, ktoré by malo umožniť používateľovi grafickú vizualizáciu vytvorenej klasterizácie.

II. METÓDY ANALÝZY PDF MALVÉRU

A. Metódy statickej analýzy

Počas statickej analýzy analytik skúma potenciálne nebezpečný súbor bez jeho vykonania. Zameriavajú sa na štruktúru a metadáta uložené v dokumente. Cieľom je zistiť, či súbor je škodlivý bez toho, aby bol spustený kód, v ňom zabudovaný. Jedna z najbežnejších a základných techník analýzy je hľadať kľúčové slová, ktoré môžu naznačovať škodlivú prítomnosť. Analytik takéhoto malvéru nemusí dekodovať alebo dekomprimovať dokument PDF, ale prehľadáva iba reťazce prítomné v dokumente na identifikáciu kľúčových slov. Analytici hľadajú kľúčové slová ako, JavaScript, OpenAction, GoTo, URI a RichMedia, ktoré naznačujú škodlivosť. Ak kľúčové slová nie sú k dispozícii, analytik môže s istotou povedať, že analyzovaný súbor je bezpečný.



Obr. 1 Prehľad metód analýzy PDF dokumentov

B. Metódy dynamickej analýzy

Metódy dynamickej analýzy využité na rozpoznanie škodlivých dokumentov sa môžu deliť do dvoch kategórií. Jedna z nich využíva statickú extrakciu Javascriptového kódu a druhá dynamickú. Obe však patria pod dynamickú analýzu, pretože sa pri ich realizácii využívajú prvky dynamickej extrakcie alebo dynamickej analýzy už oddeleného JavaScriptového kódu. Realizácia dynamickej analýzy nad celým dokumentom je síce náročnejšia na využité zdroje softvérové ale i hardvérové, avšak ponúka lepšie výsledky a vyššiu mieru indikácie malvéru v reálnom prostredí. Takiež je menej ovplyvnená obfuskáciou škodlivého kódu, pretože sa nevykonáva len nad menším refazcom.

III. NÁVRH DETEKcie ŠKODLIVOSTI PDF SÚBOROV NA ZÁKLADĚ KLASTERIZÁCIE

Základnou myšlienkou klasterizácie je rozdelenie rôznych objektov, dát či súborov do skupín, a teda klastrov podľa ich podobnosti. Tento úmysel si je možné všimnúť pri zjednodušenom pohľade na analýzu malvéru vo všeobecnosti. Pri každej novej vzorke dokumentu chceme zistiť to, do akej miery sa správa podobne doteraz známym škodlivým vzorkám. Takéto určenie podobnosti, a teda zaradenie, výrazne zefektívňuje detekciu škodlivého softvéru, pretože zaradenie do určitej skupiny vzoriek má okamžitú a jednoznačnú výpovednú hodnotu pre analytikov. Škála podobnosti je určená tak, že medzi dvoma objektami je maximálna podobnosť ak sa nachádzajú v rovnakom klasteri a minimálna naopak, ak sa nachádzajú v rozdielnom [1].

A. Algoritmus strojového učenia

Dôležitým faktorom pre čo najpresnejšiu analýzu pomocou klasterizácie je výber vhodného algoritmu pre strojové učenie. Tento algoritmus má za úlohu práve spracovanie získaných dát zo vzoriek malvéru, čo umožní zaradenie do správneho klastra. Existuje veľké množstvo klasterovacích algoritmov, avšak boli uvažované iba dva algoritmy s ohľadom na presnosť rýchlosť a efektívnosť v rámci veľkých dátových setov.

1) *Kmeans*: Algoritmus kmean je priamo založený na publikovanom článku [2]. Metóda zhľukovania kmeans sa zameriava na minimalizáciu priemernej štvorcovej vzdialenosti medzi bodmi, ktoré sú súčasťou toho istého klastra. Je daná hodnota k , ktorá označuje počet zhľukov, X označuje dátové body a V označuje množinu stredov. Kmeans je ľahko pochopiteľný a v porovnaní s ostatnými komplexnými algoritmi vykonáva zhľukovanie rýchlejšie. Vyžaduje si to však súbor údajov s dobre oddelenými bodmi nato, aby boli výsledky zoskupenia údajov presné.

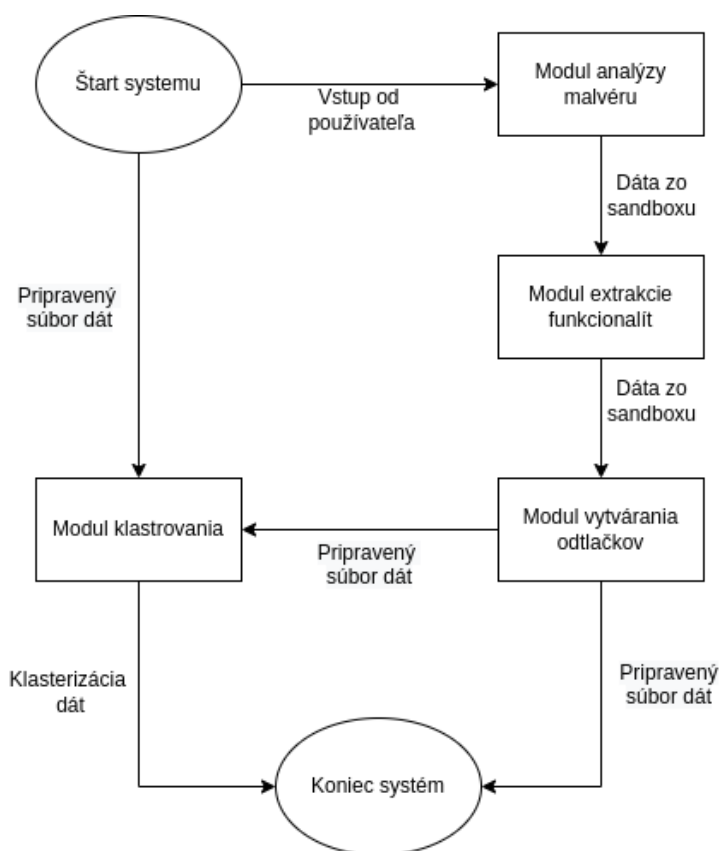
Počet zhľukov, ktoré zvolíme pre algoritmus, by nemal byť náhodný. Pre nájdenie optimálneho počtu klastrov sa využíva lakťová metóda (the elbow method). Táto predstavuje jeden z najjednoduchších spôsobov, ako získať čo najlepšiu hodnotu počtu klastrov pomocou skúmania rozptylu pri zmene počtu zhľukov. Rovnako ako mnoho iných metód vyžaduje opakované použitie klasterovacieho algoritmu. Očakáva sa, že po dosiahnutí správneho počtu zhľukov, sa percento rozptylu zníži čo indikuje najoptimálnejší počet klastrov pre daný dátový set.

2) *Selforganizing map*: Algoritmus samoorganizujúcej sa mapy (SOM) je založený na knihe [3]. SOM bol vyvinutý ako algoritmus, ktorý je súčasťou neurónových sietí, a ktorý obsahuje skrytú vrstvu. Je to veľmi zriedkavý aspekt učenia bez dozoru a je to technika vizualizácie údajov, ktorej cieľom je zmenšiť rozmery dát a odhaliť prvky. Mapy pomáhajú vizualizovať vysokorozmerné údaje. Predstavuje viacrozmerné dáta v dvojrozmernom priestore pomocou samoorganizujúcich sa neurónových sietí.

SOM sú mriežkou neurónov (alebo uzlov), ktoré reagujú na súbor vstupných signálov. Každý neurón má svoju polohu a tie, ktoré ležia blízko seba, predstavujú zhľuky s podobnými vlastnosťami. Preto každý neurón predstavuje zhľuk naučený počas tréningu dát.

Učenie prebieha v niekoľkých krokoch a v mnohých iteráciách. Vzďialenosť každého z uzlov (alebo bodu nespracovaných údajov) sa vypočíta zo všetkých výstupných uzlov. Uzol alebo neurón, ktorý je identifikovaný s najmenšou vzdialenosťou, je najlepšia zhodná jednotka (BMU). Z tohto BMU je definovaný polomer (alebo sigma). Všetky uzly, ktoré spadajú do okruhu BMU, sa aktualizujú podľa ich príslušnej vzdialenosti od BMU. [4] Po každej iterácii sa polomer znižuje a BMU sťahuje menšie uzly, takže proces sa stáva stále presnejším.

B. Konceptuálny návrh riešenia



Obr. 2 Zjednodušený konceptuálny model navrhovaného riešenia

Pre nami navrhovaný systém vzniká niekoľko základných požiadaviek na to, aby mohol byť splnený predpoklad pre úspešnú klasterizáciu malvéru. Prvým podstatným bodom je, aby riešenie bolo schopné spracovávať množinu rôznych PDF dokumentov. V našom riešení ju predstavuje dátový set obsahujúci či už škodlivé alebo neškodné dokumenty v rôznom počte, vytvorený tímom Contagio. Contagio zozbierali rôzne vzorky malvéru a voľne sprístupnili databázu viacerých datasetov rôznych typov malvéru za účelom jeho štúdie, ktoré boli odchytené pri rovnako rôznych formách útoku vrátane cielených útokov.

Ďalšou potrebnou vlastnosťou navrhovaného systému je, aby bol schopný vytvoriť odtlačok pre každý skúmaný dokument, ktorý bude systémom spracovávaný. Ten musí byť dostatočne špecifický tak, aby dokázal dostatočne rozlíšiť jednotlivé dokumenty.

Najdôležitejšou požiadavkou je, aby bolo možné naším riešením vykonať klasterizáciu PDF dokumentov. Tá by mala byť vykonávaná algoritmom Kmeans alebo algoritmom samoorganizujúcich sa máp. To nám umožní porovnať tieto dva prístupy v kontexte analýzy škodlivého softvéru,

čo je jedným z cieľov tejto práce. Požiadavkou na riešenie je vizualizácia klasterizácie pomocou jednotlivých algoritmov.

Systém, ktorého vytvorenie je cieľom tejto práce, bude delený do niekoľkých modulov. Každý z týchto modulov bude mať za úlohu splniť jednu z požiadaviek na systém ako celok. Ako prvý sa vykonáva modul analyzovania malvéru spolu s modulom extrakcie funkcionalít. Tieto moduly si vyžadujú prácu so sandboxom, pretože spracovávanie malvéru v nezabezpečenom prostredí by mohlo poškodiť systém, v ktorom navrhovaný systém beží. Tento analytický modul bude zabezpečovať okrem iného aj komunikáciu so sandboxom tretej strany. Ďalším krokom systému je spustenie modulu, ktorý zabezpečuje označenie dokumentov. Vytvára spomínaný odtlačok, ktorý má nami definovanú podobu, ktorú riešenie tejto práce podporuje. Medzi týmito modulmi je potrebné zabezpečiť prenos dát. Tie budú modulu delegované vo formáte JSON. Modul vytvárania odtlačkov naplní vopred definovaný dátový súbor špecifickým formátom odtlačkov pre každý úspešne spracovaný PDF dokument. Tento postupne napĺňaný dátový súbor predstavuje očakávaný vstup pre modul klasterizácie.

IV. IMPLEMENTÁCIA RIEŠENIA

Implementačná časť tejto práce bola vytvorená vo zvolenom jazyku Python tak, aby výsledné riešenie spĺňalo všetky požiadavky popísané v predchádzajúcej kapitole a súčasne, aby sa držalo zjednodušeného konceptuálneho modelu.

A. Modul analyzovania malvéru

Prvým, asi najzákladnejším modulom systému, je modul, v ktorom sú dokumenty, resp. malvér analyzovaný. Modul analýzy malvéru, ako už bolo spomenuté, spracováva dva druhy vstupu a teda samostatný pdf dokument alebo priečinok takýchto súborov. Pre prácu so súborami v jazyku Python sa využíva populárna knižnica Pandas. Po úspešnom načítaní súboru sa tento súbor odosiela na analýzu malvéru. Pre tento účel bol vybraný nástroj tretej strany VirusTotal. Je to online služba určená na skenovanie súborov a odkazov na vírusy. Jej hlavnou výhodou je využívanie skenerov veľkého počtu iných výrobcov. Používateľ tak môže získať objektívnejší obraz o stave a vlastnostiach svojho dokumentu. Komunikácia a odosielanie súborov na analýzu je vykonávaná prostredníctvom VirusTotal API pre jazyk Python vo forme vytvorenej voľne dostupnej knižnice *vt_py*. Pre odoslanie dokumentu na analýzu sa využíva metóda *scan_file()*, ktorá má dva parametre. Samotný súbor, na ktorom chceme vykonať analýzu a parameter *"wait_for_completion"*, ktorý ako z názvu vyplýva rieši samotné ukončenie tejto metódy. Metóda vracia objekt analýzy, ktorý je načítaný a spracovaný vo formáte JSON. Tento výsledok je pre potreby tejto práce trochu obširný, a preto je potrebné vykonať unifikáciu výsledku. Takto zredukovaný výsledok analýzy VirusTotal je poskytnutý pre spracovanie ďalším modulom na vytvorenie špecifického odtlačku dokumentu.

B. Modul extrakcie funkcionalít

Tento modul je doplnením modulu analýzy malvéru, ktorý vykoná doplnkovú statickú analýzu dokumentu a doplní výsledky modulu spúšťaného pred ním. Pre tento typ analýzy sa využívajú 2 nástroje *pdfid.py* a *pdfparser.py*, ktorých autorom je Didier Stevens [5]. Modul extrakcie funkcionalít je vlastne spojením týchto nástrojov tak, aby poskytli výsledky pre doplnenie objektu JSON. Oba tieto nástroje bolo potrebné najprv naštudovať a pochopiť ich vykonávanie a následne mierne upraviť pre potreby nášho riešenia. Taktiež bolo potrebné ponechať iba nevyhnutnú funkcionalitu. Ďalším krokom bola úprava tvorby výsledkov a ich formátu. Bolo potrebné dopísať logiku, ktorá sa stará o spracovanie výsledkov a ich transformáciu do vyžadovaného formátu.

C. Modul tvorby odtlačku dokumentov

Úlohou ďalšieho modulu je vytvorenie špecifického odtlačku pre každý analyzovaný dokument. Predpokladom pre správne spustenie tohto modulu je prítomnosť súboru *data.csv*, ktorý musí byť vytvorený v súborovom systéme projektu na najvyššej hierarchickej úrovni. Je to súbor vo formáte obyčajného textu, ktorý obsahuje zoznam údajov. Spúšťa sa v iterácii pre každý dokument nachádzajúci sa vo vstupnom dátovom sete, rovnako, ako sa odosielaajú po jednom na analýzu nástrojom VirusTotal. Modul odtlačku dokumentov je pre každý súbor spúšťaný automaticky po úspešnom dokončení predchádzajúcich modulov. Jednotlivé odtlačky sú vytvárané z hodnôt uložených v JSON výsledku predchádzajúcich modulov a tvoria ho všetky hodnoty z obdržaného výsledku.

Tvoria ho hodnoty z výsledku analýzy v poradí ako: potvrdený časový limit, zlyhanie, neškodný, škodlivý, podozrivý, časový limit, nepodporovaný typ, nezistený. Nasledujú informácie z modulu

extrakcie funkcionalít v poradí: 'obj', 'endobj', 'stream', 'endstream', 'xref', 'trailer', 'startxref', '/Page', '/Encrypt', '/JS', '/JavaScript', '/AA', '/OpenAction' a '/JBIG2Decode'.

Takto vyskladaný odtlačok súboru, by mal mať dostačujúcu formu pre potreby nášho riešenia na špecifikáciu každého skúmaného dokumentu, či obsahuje škodlivý alebo podozrivý obsah.

D. Modul klasterizácie

Modul rieši proces strojového učenia, ktoré má za úlohu vytvoriť klasterizáciu. Jeho druhou úlohou je zabezpečiť možnosť evaluácie vykonanej klasterizácie. Tento modul ponúka dva riešenia klasterizácie a to algoritmom Kmeans a taktiež algoritmom samoorganizujúcich sa máp.

1) *Algoritmus Kmeans*: Prvým krokom po tom, je potrebné načítať vstupné dáta pre tento modul. Súbor data.csv je načítavaný pomocou knižnice Pandas. Pre tento algoritmus existuje implementácia klastrovania v jazyku Python z populárneho balíka strojového učenia scikitlearn. Pomocou inštancie *cluster.KMeans* z knižnice scikitlearn je použitý kmeans s uvedením požadovaného počtu centier klastra k . [6] samotne natréovanie množiny vzoriek vstupných dát a získanie označenia príslušných klastrov je možné pomocou metódy *fit_predict(X)*.

Pre určenie optimálneho počtu klastrov sa využíva lakťová metóda. Najprv je potrebné urobiť súčet štvorcových chýb pre rôzne hodnoty K . To sme vykonali tak, že bolo spustených niekoľko kmeans, pri každej iterácii bola navýšená hodnota K a zaznamenávali sme hodnotu SSE. Následne bol použitý balík v jazyku Python s názvom kneed, ktorý na základe poskytnutých SSE dát vie určiť lakťový bod automaticky, bez skúmania grafu tejto metódy.

Samotné priradenie klastrov ku každej vzorke po vykonaní spomínanej metódy *fit_predict(X)*, je možné získať ako jednorozmerné pole NumPy z premennej *kmeans.labels_*.

Hodnoty priradených klastrov z tohto poľa sú jedným z argumentov funkcie na vykreslenie, ku ktorým sú pridané v podobe ďalšieho argumentu natréované a rozmerovo zredukované dáta. Funkcia *scatterplot()* sa používa na zobrazenie údajov a zobrazenie vzťahu medzi hodnotami pomocou vizualizácie rozptylu.

Pre možnosť evaluácie tohto algoritmu sa načítajú neupravené dáta, kde sa nachádza informácia pre každú vzorku o absolútnej ceste dokumentu, ktorému prislúcha. Postup vytvorenia inštancie *Kmeans()* a natréovania a predikcie klastrov je rovnaký ako bolo popísané vyššie. K načítaným dátam sa pridá informácia o priradenom klastru na základe indexov, aby bolo dodržané poradie pri spájaní týchto dvoch polí dát. Na základe používateľovho vstupu, ktorý obsahuje názov overovaného dokumentu, je možné získať informáciu o klastru práve pre daný dokument. Výsledný klaster overovaného súboru je vypísaný používateľovi do konzoly.

2) *Algoritmus SOM*: Po spracovaní vstupu používateľa je potrebné v prvom kroku načítať súbor data.csv, ktorý je predpokladom pre správne vykonanie algoritmu. Dáta sú načítavané bez prvého stĺpca, ktorý drží informáciu o názve dokumentu, ktorému vzorka prislúcha. Pri spracovaní týchto dát je využitá Python funkcia *iloc()*.

V jazyku Python je dostupná implementácia prostredníctvom balíka MiniSom, ktorý je bezplatný a voľne dostupný verejnosti, a tak je tento balík využívaný aj riešením tejto záverečnej práce. MiniSom je implementácia vykonaná pomocou minimalistického balíka Numpy, ktorý konvertuje vysokorozmerné údaje na nízkorozmerné údaje pre jednoduchý geometrický vzťah. [7] Po inicializácii váh nasleduje samotné tréovanie algoritmu na vstupných dátach zavolaním funkcie *som.train_random(X, iterations)*. Prvým parametrom sú naše upravené vstupné dáta, z ktorých sa vyberajú vzorky náhodne na natréovanie algoritmu. Druhým parametrom je počet iterácií, v ktorom sa má tréovanie algoritmu vykonať.

Po natréovaní je možné samoorganizujúce sa mapy graficky znázorniť. Pre vykreslenie algoritmu importujeme knižnicu pylab, ktorá sa používa na vizualizáciu výsledkov SOM a importujeme z nej rôzne balíky.

Pri spustení módu systému na overenie výsledkov pre tento algoritmus je zachovaný rovnaký postup vytvorenia inštancie *MiniSom()*, určenia váh a aj natréovania dát. Z dátového setu sa následne pri evaluácii vyberie posledná vzorka dát, ktorá predstavuje odtlačok dokumentu, ktorý je potrebné overiť. Pred overením klastra pre danú vzorku dát je potrebné transformovať poslednú vzorku z datasetu na správny formát. Na zistenie klastra, ktorému táto vzorka prislúcha, je možné použiť metódu *som.winner(x)*, ktorá vracia súradnice víťazného klastra na mape pre danú vzorku dát, ktorej dáta boli transformované. Súradnice výsledného klastra sú vypísané používateľovi v konzole.

V. VYHODNOTENIE PRÁCE

Z úvodnej časti práce vyplynula problematika klasterizácie malvéru ako tá, ktorú je možné viac preskúmať, a rovnako sa javila ako zaujímavou oblasť, ktorá spája problematiku skúmania malvéru s problematikou spracovania dát.

Tabuľka 1
Výsledky porovnania vykonávania algoritmov

	1000 vz.	2000 vz.	3000 vz.
Kmeans PC1	1.3s	1.9s	2.5s
Kmeans PC2	1.2s	1.7s	1.9s
SOM PC1	1.5s	2.1s	2.8s
SOM PC2	1.3s	1.6s	2.0s

Úmyslom použitia klasterizácie bolo overiť, či je možné zaradiť PDF dokument do klastrov na základe toho, či má znaky škodlivého obsahu alebo nie. Rovnako cieľom bolo urobiť experiment tak, aby bolo možné vybrať čo najoptimálnejšiu možnosť klasterovania, ktorá prináša čo najpresnejšie výsledky. Pre zodpovedanie týchto výskumných otázok bolo navrhnuté riešenie, ktorého cieľom bol systém na klasterizáciu dokumentov PDF, ktorý pomocou nástrojov tretích strán dokáže preskúmať ich obsah a nájsť znaky malvéru. Navrhovaný systém ponúka dve riešenia problému klasterizácie, aby bolo možné porovnať rôzne implementácie klasterovacích algoritmov. Tento celý návrh riešenia bol bližšie popísaný v kapitole návrhu riešenia.

Následne bola vyhotovená implementácia riešenia tak, aby spĺňala všetky stanovené požiadavky na navrhovaný systém.

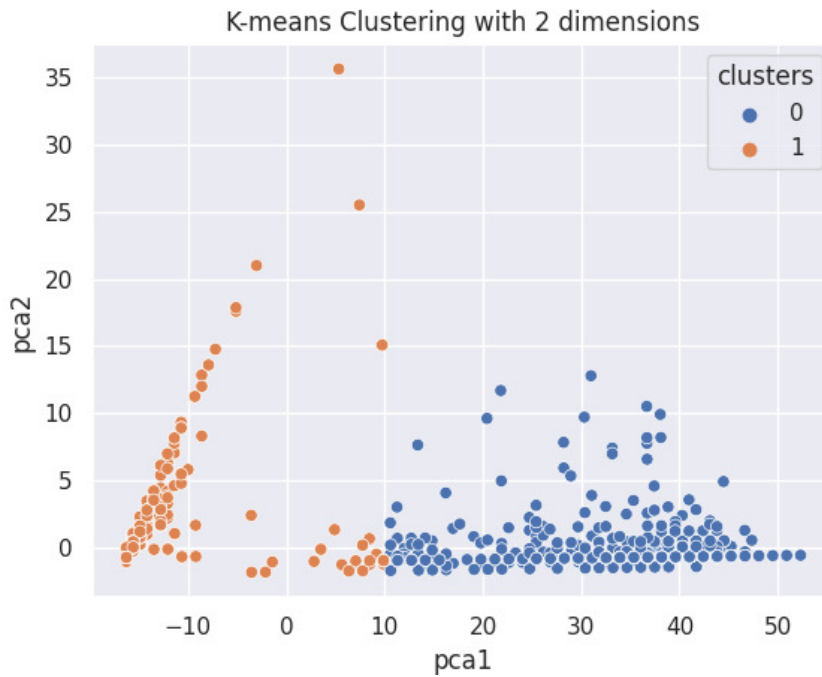
Po dokončení implementácie bolo riešenie testované na datasete dokumentov od tímu Contagio, ktorý je voľne dostupný. Dataset obsahuje takmer 12000 škodlivých a 6000 neškodných súborov. Systém bežal niekoľko dní, kedy podrobil jednotlivé dokumenty z tejto vzorky analýze a na základe jej výsledkov naplnil dataset odtlačkov, vyhotovených pre tieto dokumenty. Dĺžka prípravy datasetu trvala tak dlho z dôvodu obmedzenia bezplatného API nástroja VirusTotal, kedy mohol denne vykonať len 500 preskúmaní. Systém pri tomto testovaní bežal len v móde analýzy dokumentov, pričom boli opakovane využívané iba moduly analýzy malvéru, extrakcie funkcionalít a tvorby odtlačkov dokumentov. Výsledkom tohto testovania bolo úspešné naplnenie súboru data.csv odtlačkami pre rozličné PDF dokumenty z datasetu Contagio.

Tento súbor odtlačkov bol využitý pri následnom testovaní modulu pre klasterizáciu. Cieľom tohto testovania bolo overiť tento modul, či je pomocou neho možné vytvoriť klasterizáciu dát. Testovanie prebehlo v dvoch prostrediach. Jedno predstavoval operačný systém Windows 10 s procesorom Intel i5 a operačnou pamäťou 6GB. Druhé prostredie predstavoval operačný systém Manjaro Linux s procesorom AMD Ryzen 7 a operačnou pamäťou 16GB. Dôvodom skúšky systému v oboch prostrediach je overenie správania a podpory všetkých knížnic importovaných do Python adresára na rôznych platformách a hardvéri. Tabuľka 10.1 uvádza namerané hodnoty pri vykonávaní klasterizácie oboma algoritmi s rovnakými datasetmi o počte 1000, 2000 a 3000 vzoriek. Z tohto merania vyplýva, že algoritmus Kmeans je v zásade rýchlejší ako algoritmus SOM, aj keď rozdiely nie sú veľmi markantné. Možnou príčinou je, že algoritmus SOM vykonáva počas svojho behu množstvo iterácií pre čo najpresnejšie výsledky klasterovania. Aj týmto meraním sa potvrdilo, že Kmeans je jednoduchý algoritmus s rýchlym vykonávaním.

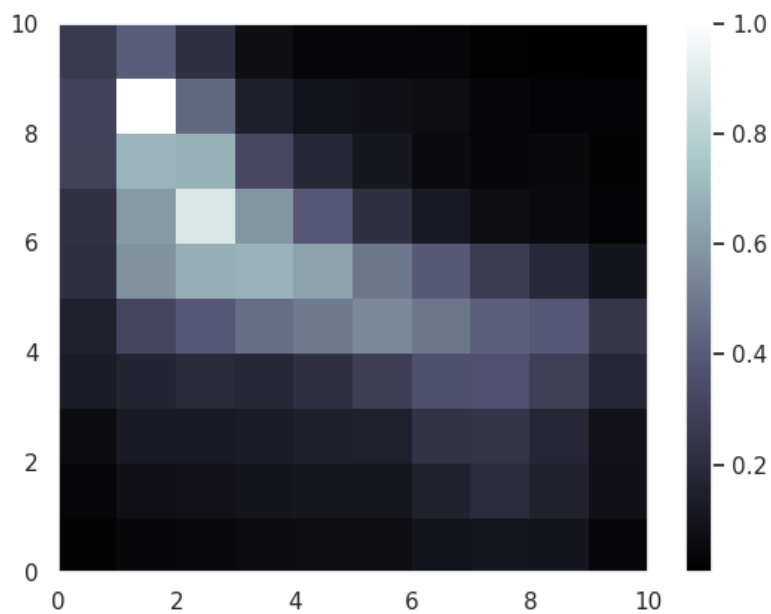
Prvým výsledkom tohto testovania bolo vizuálne zobrazenie klastrov pre oba algoritmy. Ako je vidieť na ukázkach Obr. 10.1 a 10.2, v oboch prípadoch bolo možné rozlíšiť vizuálne viacero klastrov, ktoré systém určil. Nevýhodou riešenia je, že oba algoritmy nie je možné zobrazit rovnakým vizuálnym spracovaním, čo by bolo vhodné pre porovnanie algoritmov. Možno však skonštatovať, že oba algoritmy rozdelili dátové vzorky do viacerých klastrov. V prípade tohto datasetu, klaster s označením 0 predstavuje vzorky neškodných dokumentov a cluster 1 tie škodlivé. Pre algoritmus SOM a jeho vizuálne zobrazenie sú škodlivé vzorky reprezentované klastrami s hodnotou 1 a jej blízkymi. Čím majú klaster hodnotu bližšiu k 0, čiže čiernej farbe, reprezentujú vzorky, ktoré nie sú podozrivými. Oba grafy reprezentujú rovnaký dataset s 2000 vzorkami.

Druhým výsledkom testovania tohto modulu, bolo overenie zaradenia dokumentu do niektorého z klastrov. Na základe takéhoto priradenia je možné určiť, či je analyzovaný dokument škodlivý, alebo patrí medzi dokumenty s neškodným obsahom. Pre takéto určenie je však potrebné mať vedomosť o tom, ktorý klaster predstavuje škodlivé a ktorý naopak neškodné vzorky. Je tomu tak preto, že algoritmus klasterovania nerozlišuje, čo dáta predstavujú, ale tieto dáta iba spracováva a rozdeľuje na základe podobnosti. Z toho vyplýva, že systém neodpovie používateľovi priamo, či je dokument neškodný alebo škodlivý. Je to však z jeho výsledku možné vyhodnotiť.

Táto práca má niekoľko prínosov pre oblasť skúmania malvéru a neškodného softvéru vo vybranej doméne, ktorou bol formát dokumentov PDF. Pri študovaní literatúry, týkajúcej sa tejto problematiky vo vybranej doméne však chýbali možnosti skúmania obsahu dokumentov pomocou



Obr. 3 Grafické zobrazenie klastrov pomocou Kmeans algoritmu



Obr. 4 Grafické zobrazenie klastrov pomocou SOM algoritmu

ich klasterizácie. Aj to bol jeden z dôvodov výberu tohto prístupu. Táto práca popisuje znalosti, postup, nástroje a metódy, pomocou ktorých možno takýto systém vytvoriť. Na túto prácu je možné v budúcnosti nadväzovať a vychádzať z krokov či návrhu systému. Práca dokazuje využiteľnosť klasterovania pre skúmanie malvéru aj vo formátoch ako je dokument PDF. Práca taktiež popisuje navrhnutý formát odtlačku pre skúmané dokumenty, ktorý sa vyžaduje pre úspešne vykonanie klasterizácie.

VI. ZÁVER

Prvá časť práce sa zaoberá analýzou a popisom existujúcich prístupov ku skúmaniu malvéru a ich postupov, ktoré využívajú. Na základe tejto analytickej časti bola vybraná metóda klasterizácie

malvéru, ako primárna pre skúmanie škodlivého softvéru pomocou navrhovaného systému.

V ďalšej časti práce bol vytvorený návrh systému, ktorý pozostáva z viacerých modulov. Každý z navrhovaných modulov zaobstarával naplnenie jednej, alebo viacerých požiadaviek na systém. Navrhnutý systém sa skladá z modulu analýzy malvéru, modulu extrakcie funkcionalít, modulu tvorby odtlačku dokumentov a modulu klasterizácie. V časti návrhu boli taktiež detailne vysvetlené dva algoritmy pre klasterizáciu dát, ktoré navrhovaný systém využíva.

Následne bola vytvorená implementácia, ktorá vychádzala z kapitoly návrhu. Kapitola popisujúca implementáciu riešenia zachytáva postup, ktorým sa podarilo naplniť požiadavky stanovené na systém pri návrhu riešenia. Pre niektoré požiadavky bolo možné využiť nástroje tretích strán, ako je VirusTotal pre analýzu malvéru pomocou sandboxu. Pre iné bolo potrebné vytvorenie vlastného princípu, ako v prípade tvorby odtlačku dokumentov.

Táto práca prináša niekoľko prínosov do odvetvia skúmania malvéru, predovšetkým v doméne dokumentov PDF. Hlavným prínosom je uplatnenie metódy klastrovania malvéru v týchto súboroch, čím táto práca dopĺňa škálu publikovaných prác, ktoré túto metódu neaplikovali.

Ďalšou alternatívou je možnosť v budúcnosti upraviť na implementácii model analýzy malvéru tak, aby využíval spomínaný nástroj Cuckoo sandbox. Následne by iste bolo zaujímavé previesť komparáciu výsledkov takto upraveného systému s výsledkami aktuálnej implementácie riešenia. Značným vylepšením systému by bolo taktiež použitie neobmedzenej verzie API, ktoré by ovplyvnilo najmä výkon a efektívnosť riešenia. Zaujímavé potenciálne rozšírenie práce vidíme tiež v možnosti viac sa zamerať na behaviorálnu analýzu a vykonávať klasterizáciu s cieľom zaradenia vzoriek do konkrétnych rodín malvéru.

POĎAKOVANIE

Článok bol podporený v rámci projektu 002TUKE4/2021 Implementácia moderných metód a foriem výučby v oblasti kyberbezpečnosti k požiadavkám praxe.

LITERATÚRA

- [1] M. Elingiusti, L. Aniello, L. Querzoni, and R. Baldoni, "Malware detection: A survey and taxonomy of current techniques," *Cyber threat intelligence*, pp. 169191, 2018.
- [2] J. A. Hartigan and M. A. Wong, "Algorithm as 136: A kmeans clustering algorithm," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 28, no. 1, pp. 100108, 1979. [Online]. Available: <http://www.jstor.org/stable/2346830>
- [3] T. Kohonen, *Selforganization and associative memory*. Springer Science & Business Media, 2012, vol. 8.
- [4] J. Vesanto and E. Alhoniemi, "Clustering of the selforganizing map," *IEEE Transactions on neural networks*, vol. 11, no. 3, pp. 586600, 2000.
- [5] D. Stevens, "Pdf tools," *Didier Stevens Labs*, 2008. [Online]. Available: <https://blog.didierstevens.com/programs/pdftools/>
- [6] O. Kramer, "Scikitlearn," in *Machine learning for evolution strategies*. Springer, 2016, pp. 4553.
- [7] G. U. Kiran and D. Vasumathi, "Predicting parkinson's disease using extreme learning measure and principal component analysis based mini som," *Annals of the Romanian Society for Cell Biology*, pp. 1609916111, 2021.

Tool for generating UI schemes

Jan GALOVIC, Martin HASIN, Martin HAVRILLA

Department of Computers and Informatics, Faculty of Electrical Engineering and Informatics Technical University of Košice, Slovak Republic

jan.galovic@student.tuke.sk, martin.hasin@tuke.sk, martin.havrilla@tuke.sk

Abstract – The master thesis deals with the issue of creating schemas for forms and creating tool for generating UI schemas using Drag and Drop in React. It focuses on current ways of creating schemas for forms, introduces the individual formats for these schemas and describes their use. It also introduces existing tools for creating UI schemas and working with JSON schemas. Subsequently, it focuses on the creation of a tool that can create a JSON UI schema based on the essential elements in the form. This tool has actually been used in the design and generation of UI schemes, where it has been shown to be beneficial and simplifies the work of creating UI schemes.

Keywords – UI scheme, React, Drag and Drop, react-beautiful-dnd, JSON scheme, Form, Generator UI

I. INTRODUCTION

Nowadays, the user environment plays a very important role in systems or applications. To interact with the system, the user needs to enter, edit, or view data in the environment. Whether it's entering data to send an order when shopping online, registering in a system or application, or just filling out a questionnaire, it's always some representation of the data in the form. In order for this data to be easy to read and easily modify by the user, it is necessary to adhere to UX principles, which speak of the correct grouping and display of elements in a given form. As everything is currently handled with mobile devices and web applications, responsive design is required. At the same time, it is necessary to ensure the simplicity of the form without disturbing elements that could discourage the user from reusing the application.

The ways in which these forms can be created are constantly expanding, with almost every programming language having a library for creating form elements. Examples are "WinForm" for Windows applications, "Xamarin.Forms" for mobile applications or Bootstrap for web applications. This is most often represented in web compilers using HTML attributes and CSS properties together with the Bootstrap library. The individual elements of the form can then be represented in the form of text, numbers, check boxes, scrolling and many other elements that can be displayed.

II. STANDARDS FOR CREATING FORMS

Forms are one of the inseparable types of user interactions on the web, but also in applications. As Jinesh Darji writes in his article [1], each user who accesses a website or application has a specific goal. This goal can be different, whether it is an online purchase, registration in the system, or just filling out a questionnaire. There is always a form between the user and this goal. Each form has a purpose and is expected to be completed by the user so that the correct data is collected as a result. In order for this data to be filled in correctly, it is necessary to follow the practical principles that have been proposed from testing in practice, through eye tracking [2], usability testing or observations and comments directly from users. These principles consist of:

- Form structure - correct sorting and grouping of data
- Defining input fields - mandatory and optional fields, autofocus
- Field designation and naming - a unique name for the field
- Action buttons - visible division of the primary and secondary button

III. CREATING FORMS AND DATA DISPLAY IN THEM

There are many ways to create or write a schema for forms in different systems or programming languages. We will mention only some of them and we will deal with only a specific type for our tool for generating the UI schema.

A. Web Forms, HTML and CSS

The primary user input on the web is HTML web forms, which are used for a variety of reasons, including registration, online store sales, or security. Sometimes the process of filling out these web forms can be complicated. The first web applications focused more on simpler tasks, such as finding and browsing large amounts of data, but nowadays people use the web to perform complex intercompany and internal processes [1]. Although programmers try to ensure that processes are as simple to use and easy to understand as possible, users often fill out the basic form incorrectly. It is critical that the user understands how to complete the form correctly, as misunderstanding or incorrect completion of the form can lead to a loss of traffic to the site and thus to a loss of income [2]. Therefore, it is very important that we use the correct styling using CSS and the placement of elements in the form, so we ensure better clarity and thus better data retrieval, as it will reduce the error rate on the part of the user.

HTML solves creating forms using the pair element `<form>`. This element contains additional `input` attributes, which vary depending on what we need to display in the form. The basic input types are:

- text - Displays basic one-line text input
- radio - Allows you to select one of several options
- checkbox - Allows you to choose from several options from none to all
- submit - Displays a button for submitting the form
- button - Displays the button to which the action can be set

B. JSON forms, JSON objects and JSON data schemas

JSON Form is a declarative [3] framework for efficiently creating form-based Web user interfaces. These user interfaces are focused on entering, editing and viewing data and are usually embedded in the application. Handwriting HTML templates and Javascript for data binding is difficult, especially in reasonably sized applications. In addition, the form is often more than just a set of input fields and more advanced functionality is required, e.g. validation or conditional visibility. The JSON form uses the capabilities of JSON and the JSON schema and provides a simple and declarative way to describe forms. The forms are then rendered using a library or user interface framework, e.g. React or Angular. Any UI user interface is defined using two schemes:

- Data, JSON schema, defines the basic data to be displayed in the user interface (objects, properties and their types)
- The UI schema defines how this data is rendered as a form, e.g. the order of the controls, their visibility and layout.

Both of these schemas are interpreted by the framework at runtime and mapped to the appropriate components of the user interface, which includes data bindings, validations, and so on.

JSON Forms Architecture

JSON forms are based on the main module `jsonforms/core` [4], which provides tools for managing and rendering forms based on the JSON schema. The basic package is independent of any user interface technology.

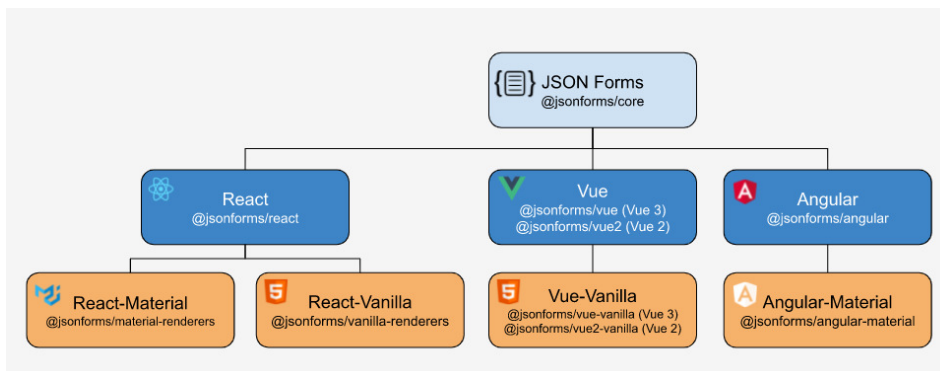


Fig. 1 JSON Forms Architecture

The following modules are: JSON Forms React `jsonforms/react`, JSON Forms Angular `jsonforms/angular` and JSON Forms Vue `jsonforms/vue`, `jsonforms/vue2`. These modules use the basic

package to provide specialized bindings for React, Angular and Vue. This approach is especially useful when developing multiple rendering sets over the same technology (i.e., React), as basic bindings do not need to be reimplemented with each set. There are two other sets of renderers for React: `jsonforms/material-renderers`, which is based on the popular Material-UI framework, and `jsonforms/vanilla-renderers`, which provide pure HTML5 renderers. For Angular, it is a set of renderers based on Angular Material `jsonforms/angular-material`. For Vue, it's a rendering set based on HTML5, `jsonforms/vue-vanilla` for Vue 3, and `jsonforms/vue2-vanilla` for Vue 2.

IV. TOOL DESIGN FOR WORK WITH JSON UI SCHEMA

Designing your own tool to generate different schemes is not as trivial as it might seem at first glance. Therefore, our goal is to create a tool that would simplify the work of a programmer or designer in creating a JSON UI schema. This tool should be clear and intuitive when creating more complex schemas for different forms, which would otherwise require more time to design the layout of individual elements in a given form. At the same time, we want the user to be able to easily and quickly manipulate individual elements during deployment and not have the knowledge needed to manually write JSON UI schemas, where errors can occur. This tool itself will generate this schema based on the distributed elements in the worksheet. These elements in the form are entered by the user and distributed by moving them in this environment according to predefined specifications and UX principles.

A. Programming environment and programming language selection

We implement the practical part in the programming environment, Visual Studio Code, with installed extensions "GitHub Pull request and Issues" and "ES7 + React/Redux/React-Native snippets". As a programming language, we chose JavaScript with the React front-end library, which is ideal for creating user interfaces or user interface components. It is also useful to use this library as a basis for developing single-page applications, so it was an ideal choice for creating our tool. It is also important to mention that the basic components of `react-web-components`, which are part of the projects at CoraGEO and help create the application itself, have been imported because they contain already programmed components that can be used without further modification.

V. IMPLEMENTATION FOR GENERATING UI SCHEMA

After creating the project and importing the basic libraries and elements that have already been created, we have tried to create only a simple tool that would display the JSON schema, UI schema, data schema and the resulting form. This part of our tool will present how the individual diagrams work with each other and what result they can display together. In the following chapters, we discussed in more detail how we proceeded and what the final result of this section was.

A. JSON, UI and data schema

The initial phase was to prepare individual diagrams, which we will later display in text boxes. These schemes will only be helpful and can be further modified directly by the user to see the resulting form. This form will be displayed using the Detail component, which takes as input parameters three schemas, namely a common schema, a UI schema and a data schema. The component itself also contains other parameters, which can be used to additionally set the remaining parameters.

B. Using the Detail component

The necessary JSON schemas are ready so we can proceed to view the form itself with these schemas. Since the goal is to allow the user to change these diagrams and really see the result immediately, it was necessary to divide the screen into 4 parts where in three parts there will be text fields, Textarea component containing the diagrams and the last part will represent the resulting form, Detail component.

This component contains a number of parameters that we can set and adapt to our requirements. For us, the most important are only the schema, ui, data with which we enter inputs from the current state in the application. The initial state after the start can be seen in the following picture



Fig. 2 The resulting layout of components

VI. IMPLEMENTATION "DRAG AND DROP" FOR GENERATING A UI SCHEMA

To accomplish this part of our work, it was necessary to create a component that we could use to move and place form elements. It is more convenient for the user to enter the set of elements he requires within the form at the beginning and then place them according to his ideas. It is best to use a drag and drop library to move elements. Since there are many libraries that allow drag and drop interactions within React, the primary task was to examine the individual libraries that allow this interaction. The main library is react-dnd [5], which contains a large set of primitive elements that work very well with the inconsistent drag and drop function in html5. It contains basic elements for the need to move elements within one sheet, there is also documentation for this library to describe the functionality and procedures of use. For the purposes of this work, the most suitable was the library, react-beautiful-dnd [6], which satisfies higher-level abstractions specifically for lists with vertical or horizontal movement, even within nested lists.

A. Component with react-beautiful-dnd library

After creating the component, to which we will add the necessary functionality for the interaction of elements using drag and drop, we need to install the library of our choice. This installation was performed using the command "npm install react-beautiful-dnd --save". It was also necessary to import the necessary functionality for the component:

- DragDropContext - represents the packaging of the part of the application for which we want to allow moving elements with the mouse
- Droppable - is an area where the element to be moved can be inserted and contains Draggable
- Draggable - represents elements that can be moved

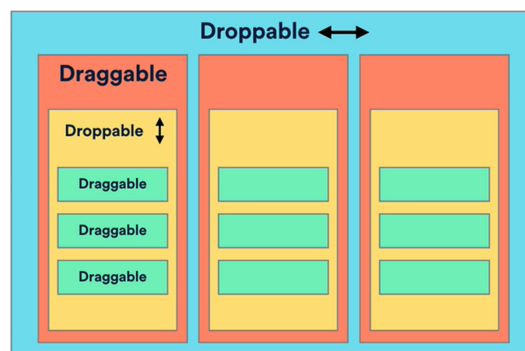


Fig. 3 Droppable and Draggable nesting illustration

B. Components - Row, Column and Item

The main component of the Generator consists of several nested components. This division was chosen for better clarity and the ability to assign functionality and parameters specifically for a given level in the form. Since I need to ensure the nesting of three levels, row, column and element, we have also oriented the naming of individual nested components. Within these components, it was also necessary to add or remove an element of the appropriate level. The individual components also had to be wrapped in "Droppable" and their interior in "Draggable" to allow the individual elements to be moved correctly at the appropriate level.

C. Functionality of moving elements in a matrix

In order to be able to move elements within the form, it was necessary to add functionality that will take care of the correct handling of these elements. The element moving functionality was resolved first. Because the elements are located in individual columns, it was first necessary to resolve moving the element within a single sheet in the column. Here it was necessary to define what element and from what position it was taken and in which position it was placed. Subsequently, the status in the form was updated.

```
if (start === finish) {  
  const newItemIds = Array.from(start.itemIds);  
  newItemIds.splice(source.index, 1);  
  newItemIds.splice(destination.index, 0, draggableId);  
  
  const newColumn = {  
    ...start,  
    itemIds: newItemIds,  
  };  
  
  this.setState({  
    columns: this.state.columns.map((col) => {  
      if (col.id === destination.droppableId) {  
        return newColumn  
      }  
      return col;  
    })  
  });  
  return;  
}
```

Fig. 4 Moving elements in List

VII. USING A GENERATOR IN PRACTICE

From the database script of the given table it was necessary to list the individual elements directly into the generator, in which it is possible to further work with the given elements. Since the generator allows us to move the elements, we do not have to pay attention to the order in which the elements are written in the script when rewriting, because in the following steps the elements will be moved to the desired position.

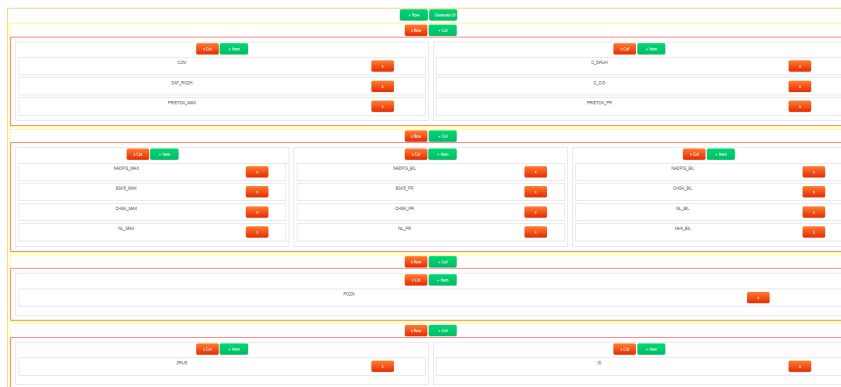


Fig. 5 Final layout of elements in the generator

The whole process of creating a JSON UI schema for a given form was simplified and accelerated just by using the tool, because the given schema did not need to be rewritten manually. The individual elements of the form only needed to be correctly defined and then grabbed and released to the selected place so that it meets the set requirements. We received the final layout of the form in a much shorter time than with the old procedure. After the final layout, the JSON UI schema was simply generated. Subsequently, it could be additionally extended by "customComponent", or even used in the database. The final form looks exactly the same, but the time was shorter and the complexity of the design less.

The form using the JSON UI schema from the generator looks exactly the same as handwriting, but such knowledge of writing the JSON UI schema is far from required.

Fig. 6 Display form after using JSON UI schema from generator

VIII. CONCLUSION

The aim of this thesis was the analysis of current standards and ways of displaying data in forms, where we got acquainted with UX rules in creating forms. We envisioned the importance of following these rules for both web and mobile applications. We also introduced various libraries that are used to create forms across various platforms, from Windows applications to web applications to mobile applications. The subject of the research was also the analysis of the most used JSON schemes, which provide a simple and declarative description of forms, which ensures validation and functionality over individual elements of the form.

In the practical part of the thesis was the design and implementation of its own tool for working with JSON schemas and generating JSON UI schemas, in a simple and efficient way, so that we avoid errors in manually defining this matrix. A part was also included in the design where it was possible to easily simulate how the JSON scheme, UI scheme and data scheme work with each other. Subsequently, we used the tool we created in practice directly during the transition of the old system to the new RETEC technology, where it was necessary to redefine the layout of elements in the forms. The use of the tool has proven to be an effective tool, as it has greatly speeded up the work of generating JSON UI schemas that would otherwise have to be written by hand. This tool was used at work by several programmers, who were also satisfied with the efficiency and ease of use. The first part of the tool, which served as an example of how the individual JSON schemes work and cooperate, was also beneficial. This was especially beneficial for people who switched from the old technology, and the use of the REACT and REDUX frameworks together with JSON schemes was new to them.

ACKNOWLEDGMENT

This publication has been published with the support of the Operational Program Integrated Infrastructure within project: Research in the SANET Network and Possibilities of Its Further Use and Development (ITMS code: 313011W938), co-financed by the ERDF. Development and innovation of TUKE simulation environment in the field of technical sciences (004TUKE-2-1 / 2021)

REFERENCES

- [1] J. Darji, "A comprehensive guideline for designing most efficient web forms: On structure, input fields, labels and action buttons," Bacancy technology, Tech. Rep. 2, 4 2020. [Online]. Available: <https://www.bacancytechnology.com/blog/standard-practices-for-designing-most-efficient-web-forms>
- [2] T. Nomura and Y. Mitsukura, "Analytical methods for web form usability," in 2015 10th Asian Control Conference (ASCC), 2015, pp. 1–6.
- [3] EclipseSource, UI Schema Elements, EclipseSource, 2021, manuals for Using JSON Forms. [Online]. Available: <https://jsonforms.io/docs/uiscema/uiscema>
- [4] —, Architecture, EclipseSource, 2021, manuals for Using JSON Forms. [Online]. Available: <https://jsonforms.io/docs/architecture/>
- [5] Drag and Drop for React, React DnD, MIT, 2022, documentation for React DnD, Drag and Drop for React. [Online]. Available: <https://react-dnd.github.io/react-dnd/docs/overview>
- [6] React beautiful dnd, Atlassian, 2022, beautiful and Accessible Drag and Drop with react-beautiful-dnd. [Online]. Available: <https://github.com/atlassian/react-beautiful-dnd>

Koncentrácia pevných častíc v ovzduší: porovnanie meraní

¹Simona KIREŠOVÁ, ²Milan GUZAN

^{1,2} Katedra teoretickej a priemyselnej elektrotechniky, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

¹simona.kiresova@tuke.sk, ²milan.guzan@tuke.sk

Abstrakt — V článku je aj na základe negatívnych skúseností s inými typmi senzorov predstavené overenie presnosti senzorov merajúcich pevné častice v ovzduší s nasledovným meraním na dvoch rôznych miestach. Výsledky meraní v tom istom čase dvomi senzormi SPS30 potvrdili zhodu s údajmi uvedených v katalógovom liste výrobcu a merania na dvoch rôznych miestach potvrdili horšiu kvalitu ovzdušia v blízkosti domu v porovnaní s meraním na záhrade.

Kľúčové slová — ovzdušie, pevné častice, senzory, SPS30

I. ÚVOD

Rôzne typy senzorov merajú s rôznymi presnosťami. V rámci tohto istého typu je výrobcom pre senzor zaručovaná určitá presnosť. Pri rovnakých typoch senzorov sa jednotlivé senzory budú určite svojimi nameranými hodnotami líšiť, ale rozptyl nameraných hodnôt by mal byť v rámci presnosti senzora. Pri potrebe merania koncentrácie pevných častíc v ovzduší je potrebné vedieť, či používané senzory merajú v rámci výrobcom deklarovanej presnosti. Vzhľadom na nie dobré skúsenosti s inými typmi senzorov bude pred samotným nasadením oboch senzorov zaujímavé zistiť, ako sa budú vzájomne líšiť ich namerané hodnoty.

II. PRESNOSŤ VIACERÝCH SENZOROV ROVNAKÉHO TYPU

Na vlastnosti senzorov vplývajú viaceré faktory, akými sú napr. podmienky skladovania senzora pokiaľ je senzor mimo prevádzky. Výrobcovia zvyčajne uvádzajú minimálnu a maximálnu hodnotu skladovacej teploty a vlhkosti. V závislosti od typu senzorov je niekedy uvažovaná aj prítomnosť niektorých plynov či maximálny tlak. Vplyv teploty pri prevádzke musí byť tiež zohľadnený. Rozsah prevádzkových teplôt je daný dolnou a hornou hraničnou teplotou, pri ktorej si snímač zachováva svoju presnosť. Mnohé charakteristiky senzorov sa menia s teplotou, a tak výrobcovia častokrát uvádzajú v katalógoch presnosť pomocou grafickej závislosti, alebo v tabuľke uvedú presnosť pre viacero rozsahov teplôt. Teplota ovplyvňuje aj dynamické vlastnosti, napríklad rýchla zmena teploty môže spôsobiť generovanie rušivého signálu na výstupe senzora [1]. Súčasťou špecifikácie presnosti je aj krátkodobá a dlhodobá stabilita. Kým krátkodobá stabilita sa prejavuje ako zmena výkonu senzora v rámci minút, hodín, resp. dní; dlhodobá stabilita je spojená s degradáciou materiálov, z ktorých je snímač vyrobený, a teda je uvažovaná z dlhšieho časového obdobia, ktorým sú mesiace až roky [1].

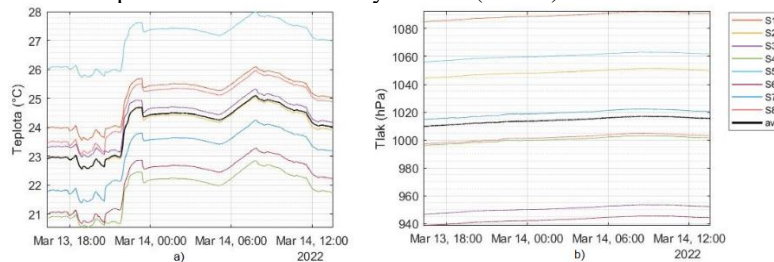
Avšak, do akej miery senzory aj reálne spĺňajú katalógové vlastnosti udávané výrobcom, môže byť v určitých prípadoch otázne. Ako príklad môžeme uviesť meranie teploty a atmosférického tlaku pomocou ôsmich kusov senzora BMP280. Senzor pracuje na princípe piezoelektrického javu, pri ktorom vplyvom mechanického zaťaženia tlakom dochádza k deformácii kryštálu, čím vzniká elektrické napätie. Senzor BMP280 komunikuje s mikrokontrolérmi resp. vývojovými doskami pomocou rozhraní I2C alebo SPI. Tlak meria v rozsahu 300 až 1100 hPa a teplotu v rozsahu -40 až 85 °C. Jeho presnosť je uvedená v Tab. 1 [2].

Tab. 1
Presnosť senzora BME280 [2].

Veličina	Typická presnosť
Tlak pri: -20 až 0 °C → 0 až 65 °C	± 1,7 hPa → ± 1,0 hPa
Teplota pri: 25 °C → 0 až 65 °C	± 0,5 °C → ± 1,0 °C

Obr. 1 ilustruje zmeny teploty a tlaku pomocou ôsmych senzorov BMP280. Z Obr. 1a), na ktorom je graficky znázornený priebeh nameranej teploty v čase, je možné vidieť, že rozptyl nameraných teplôt dosahuje až 5 °C, čo nevyhovuje údajom uvádzaných výrobcom, ktorý pri teplote 0 až 65 °C zaručuje presnosť ± 1 °C a pri teplote 25 °C je uvedená presnosť dokonca iba $\pm 0,5$ °C.

Obr. 1b) znázorňuje priebeh nameraného tlaku v čase, pričom rozptyl tlaku je až 60 hPa, čo nie je porovnateľné s presnosťou zaručenou výrobcom (Tab. 1).



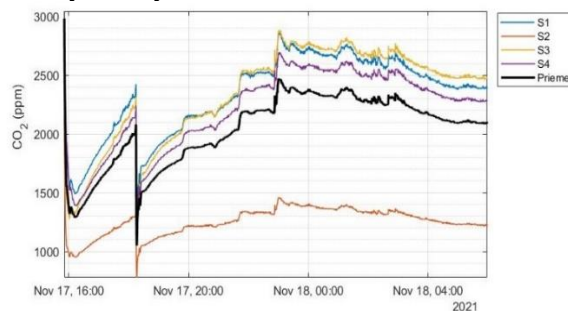
Obr. 1 Meranie a) teploty a b) tlaku ôsmymi senzormi typu BMP280.

Cieľom merania viacerými senzormi bolo zvýšiť presnosť, čo dosiahneme výpočtom priemernej hodnoty z výstupu niekoľkých senzorov, ktoré snímajú tú istú veličinu za rovnakých podmienok a v rovnakom čase. Chyba je znížená o druhú odmocninu z počtu senzorov použitých v meracom systéme. Teda po výpočte priemeru z ôsmich senzorov očakávame chybu $1/\sqrt{8}$, teda 0,35355 °C (resp. hPa) [3]. To platí iba za predpokladu, že jednotlivé senzory použité v meracom systéme odpovedajú presnosti stanovenej výrobcom, čo neodpovedalo skutočnosti pri meraní teploty a tlaku na Obr. 1.

Čo teda mohlo spôsobiť vyššiu nepresnosť senzorov BMP280? Ako príčinu môžeme vylúčiť starnutie senzorov a stratu požadovaných vlastností v rámci ich dlhodobej stability, nakoľko Obr. 1 zobrazuje prvé meranie senzorov po ich kúpe. Vylúčiť môžeme aj zmenu vlastností v rámci krátkodobej stability, keďže aj po niekoľkých hodinách ostáva rozptyl nameraných hodnôt rovnaký ako na začiatku merania. Dokonca aj ďalšie merania počas nasledujúcich dvoch mesiacov vykazovali podobné výsledky. Teda ako príčinou nepostačujúcej presnosti senzorov BMP280 ostáva buď chyba pri výrobe, alebo nevyhovujúce podmienky uskladnenia pred kúpou senzoru či mechanické poškodenie senzora pri jeho preprave. Je potrebné však podotknúť, že výroba senzora BMP280 bola ukončená, čo poukazuje na to, že najpravdepodobnejšou príčinou nepostačujúcej presnosti z vyššie uvedených môže byť práve chyba pri výrobe. Firma BOSCH, ktorá senzor BMP280 vyrábala, má však v ponuke zaradené novšie typy senzorov, ako sú napríklad senzory teploty a tlaku BMP384 a BMP388 [4], [5].

Uvádzanie zaručenej presnosti v katalógu je dobrou praktikou, avšak nie každý výrobca ju akceptuje. Dostávame sa takto napr. k senzoru kvality ovzdušia MQ-135, ktorý vo svojom katalógu uvedenú presnosť nemá. Citlivá časť senzora je pokrytá vrstvou oxidu ciničitého (SnO_2), ktorá mení svoj odpor v závislosti od koncentrácie plynov, ako je napr. CO_2 (oxid uhličitý), NO_x (oxidy dusíka), NH_3 (amoniak), alkohol, benzén a pod. V katalógu je uvedené, že pred prvým použitím senzora je pre správnu funkčnosť potrebné senzor predhriať aspoň na 24 hodín [6].

Na Obr. 2 sú graficky znázornené výsledky merania CO_2 pomocou štyroch senzorov MQ135. Zatiaľ čo rozdiel medzi senzorom S3 (žltá čiara) a senzorom S1 (modrá čiara) je najviac 150 – 200 ppm, rozdiel medzi senzorom S3 (žltá čiara) a senzorom S4 (fialová čiara) je približne dvojnásobný (300 – 400 ppm). Najviac problematickým je však senzor S2 (červená čiara), ktorého rozdiel od senzora S3 sa pohybuje v rozmedzí 1000 – 1500 ppm, čo je dvojnásobok nameranej koncentrácie CO_2 senzorom S2. Pretože výrobca neuvádza žiadnu zaručenú presnosť, nevieme, aký rozptyl medzi senzormi máme očakávať. Rozdiel medzi senzormi takmer 1500 ppm pri maximálnej nameranej maximálnej hodnote takmer 2900 ppm je príliš veľký na to, aby bol aj tento senzor v budúcnosti využívaný.



Obr. 2 Meranie oxidu uhličitého štyrmi senzormi MQ135.

Vyššie opísané skúsenosti s používaním dvoch nevyhovujúcich senzorov sú však ojedinelé, pretože pri meraní s viacerými senzormi SHT31 SMART GADGET, Dallas DS 18B20+ [7], [8], DHT11, SHT31 či SHT85 [9], bol rozptyl nameraných hodnôt v tolerancii uvádzanej výrobcom. V súčasnosti je aktuálne sledovať a vyhodnocovať koncentráciu pevných častíc (angl. particulate matter (PM)) v ovzduší. Dostupným senzorom je SPS30 od firmy Sensirion, ktorej senzory teploty SHT31 SMART GADGET, SHT31 či SHT85 merajú v rámci presnosti zaručenej od výrobcu. Prečo je dobré venovať pozornosť meraniu PM práve v súčasnosti a aké budú rozdiely v nameraných hodnotách medzi oboma senzormi?

III. MERANIE PEVNÝCH ČASTÍC SENZORMI SPS30

Prítomnosť PM v ovzduší má negatívne vplyvy na zdravie človeka, čo je dôvodom na monitorovanie ich koncentrácie. Pevné častice negatívne ovplyvňujú dýchaciu a kardiovaskulárnu sústavu. Spôsobujú zníženú funkciu pľúc, čo spôsobuje ťažkosti pri dýchaní, astmu, nepravidelný krvný tep či srdcový infarkt [10]-[13]. Čím sú PM menšie, tým väčší dopad majú na zdravie, keďže menšie častice sú schopné preniknúť hlbšie do dýchacej sústavy. Kým hrubé častice sa usádzajú v hornej dýchacej sústave, jemné častice môžu preniknúť až do pľúcnych alveol a ultrajemné častice sa môžu dostať až do krvného obehu [10]. Odhaduje sa, že až 5 % úmrtí v dôsledku rakoviny pľúc je spôsobených tým, že je organizmus vystavený PM. Tie ovplyvňujú zdravie ľudí vo všetkých vekových skupinách, avšak najviac náchylnými voči ich vplyvu sú deti, seniori a tehotné ženy [14]-[20].

Kvôli vyššie uvedeným skutočnostiam teda považujeme za meranie PM v ovzduší za aktuálne. Dôležité pri ich meraní je vedieť, či senzory, s ktorými PM meriame sú spoľahlivé a poskytujú nám správne výsledky.

Veľkosť a zloženie PM sa môže líšiť. Podľa veľkosti rozlišujeme nasledovné kategórie PM: PM10 alebo častice s aerodynamickým priemerom menším ako 10 μm ; PM2,5 (jemné častice, angl. fine particles) alebo častice s priemerom menším ako 2,5 μm ; hrubé častice (angl. coarse particles) s priemerom o veľkosti 2,5 až 10 μm a ultrajemné častice (angl. ultrafine particles) alebo PM0,5, čo sú častice s priemerom menším ako 0,5 μm [21]. Samozrejme, podľa veľkosti priemeru častíc je možné definovať aj iné kategórie PM. Napr. častice PM4 by mali priemer menší než 4 μm a PM1 by mali priemer menší než 1 μm . V praxi sú však najčastejšie merané PM spadajúce do kategórií PM10 a PM2,5 čo sa týka ich hmotnostnej koncentrácie [23]-[25]. V tomto článku sa zameriame na meranie PM10.

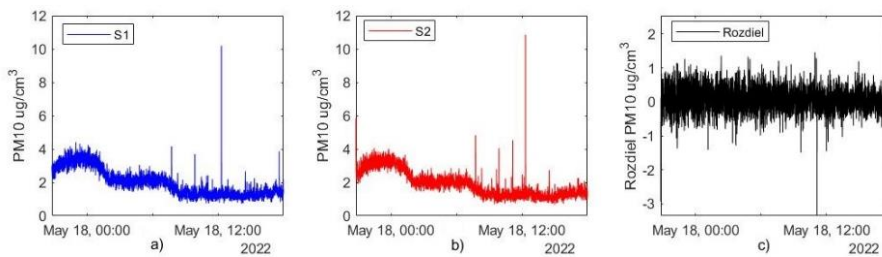
Pri našich meraniach používame senzor SPS30 vyrábaný spoločnosťou Sensirion. Ide o optický čítač častíc (angl. Optical Particle Counter (OPC)). Snímanie senzora je založené na princípe rozptylu svetla. PM sú zasiahnuté laserovým lúčom pod určitým uhlom, čo spôsobí tienenie a rozptyl svetla. Zariadenie zosníma intenzitu svetla po tom, čo bola PM zasiahnutá laserovým lúčom, čo umožňuje spočítanie a zmeranie jednotlivých častíc. Častica prechádzajúca cez osvetľovaciu zónu vytvára fotoelektrický impulz, ktorého výška závisí od veľkosti častice [26]-[29].

Senzor SPS30 je schopný zosnímať hmotnostnú koncentráciu častíc PM1, PM2,5, PM4 a PM10, ďalej početnú koncentráciu častíc PM0,5, PM1, PM2,5, PM4 a PM10 a taktiež typickú veľkosť častíc. Minimálna veľkosť PM ktoré je senzor schopný detegovať je 0,3 μm . Presnosť senzora zaručená výrobcom je v Tab. 2 [30].

Tab. 2
Presnosť senzora SPS30 [30].

Parameter	Podmienky	Presnosť
Hmotnostná koncentrácia PM1 a PM2,5 → PM4 a PM10	do 100 $\mu\text{g}/\text{cm}^3$	$\pm 10 \mu\text{g}/\text{cm}^3 \rightarrow \pm 25 \mu\text{g}/\text{cm}^3$
	nad 100 $\mu\text{g}/\text{cm}^3$	$\pm 10 \% \rightarrow \pm 25 \%$
Početná koncentrácia PM1 a PM2,5 → PM4 a PM10	do 1000 častíc/ cm^3	$\pm 100 \text{ častíc}/\text{cm}^3 \rightarrow \pm 250 \text{ častíc}/\text{cm}^3$
	nad 1000 častíc/ cm^3	$\pm 10 \% \rightarrow \pm 25 \%$

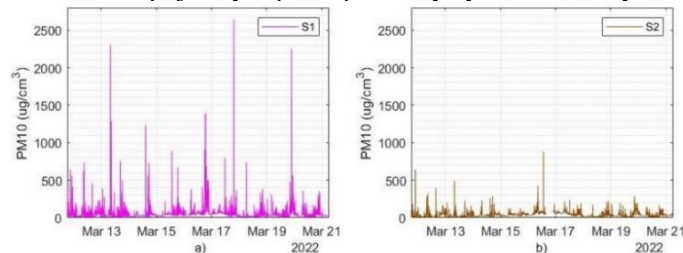
Pre meranie boli použité dva senzory SPS30. Meranie sa uskutočnilo v Košiciach na balkóne na 5. poschodí bytového domu. Obidva senzory boli umiestnené tesne vedľa seba a pripojené cez multiplexor k meracej stanici. Tá okrem senzorov SPS30 a multiplexera pozostáva z vývojovej dosky Wemos D1 R2, RTC modul, microSD modul s microSD kartou (na ktorú sa namerané hodnoty uložia každých 5s), senzora teploty a tlaku MS5611 a senzora teploty a vlhkosti SHT30. Veličiny teplota, vlhkosť a tlak sú merané za účelom zistenia možnej korelácie medzi nimi a koncentráciou pevných častíc, čo však už nebude predmetom tohto článku. Zaujímalo nás, ako veľmi sa budú odlišovať namerané hodnoty oboch senzorov v porovnaní s uvedenými katalógovými údajmi presnosti senzora. Výsledky merania ilustruje Obr. 3.



Obr. 3 Hmotnostná koncentrácia PM10 v ovzduší meraná dvomi senzormi SPS30 (a, b) a ich rozdiel (c)

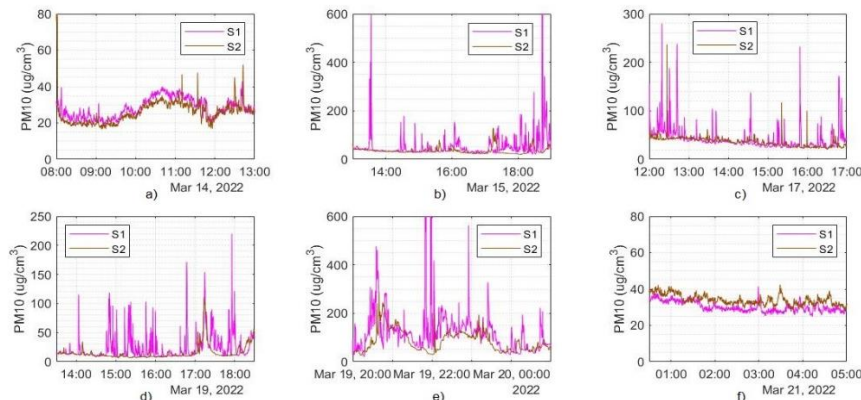
Ako je z Obr. 3 možné vidieť, priebeh koncentrácie PM10 nameraný senzorom S2 (Obr. 3b) kopíruje priebeh nameraný senzorom S1 (Obr. 3a). Z Obr. 3c), ktorý ilustruje rozdiel medzi senzormi (t.j. $S1 - S2$) je zrejme, že pre väčšinu nameraných hodnôt sa rozdiel medzi nimi udržiava v okolí nuly. Výnimkou sú úseky špičiek vyššej hmotnostnej koncentrácie PM10 (napr. pri prudkom vzostupe hmotnostnej koncentrácie PM z 2 na $12 \mu\text{g}/\text{cm}^3$ a nasledovnom prudkom poklese) ktoré ale stále odpovedajú výrobcom určenej presnosti (Tab. 2).

Ďalšie meranie, ktorého priebeh je zobrazený na Obr. 4, bolo uskutočnené v malej dedine, na balkóne rodinného domu a v záhrade. Balkón je situovaný na ulicu a na ľavej aj pravej strane sú v susedstve dva rodinné domy. Oproti cez ulicu sú ďalšie dva rodinné domy. Všetky domy využívajú drevo ako zdroj tepla. Boli zostavené dve meracie zariadenia, pričom jedno zariadenie bolo umiestnené na balkóne, 1.posch. (Obr. 4a) a druhé zariadenie bolo umiestnené na záhrade (Obr. 4b). Medzi obidvoma meracími zariadeniami bola vzdialenosť cca 35 m (najbližšie iné rodinné domy boli od miesta merania na záhrade vzdialené 40-50 m). Meranie teda bolo uskutočnené v oblasti, kde sa vo väčšine rodinných domov vykuruje spaľovaním dreva, čo je jedným z najväčších zdrojov PM [22], [31]. Účelom tohto merania bolo zistiť, do akej miery vplyva vzdialenosť od zdroja pevných častíc na ich hmotnostnú koncentráciu, pretože v čase vykurovania je na ulici cítiť nepríjemný zápach spôsobený dymom z okolitých komínov.



Obr. 4 Pevné častice PM10 merané na: a) balkóne a b) záhrade, vzdialených od seba 35 m.

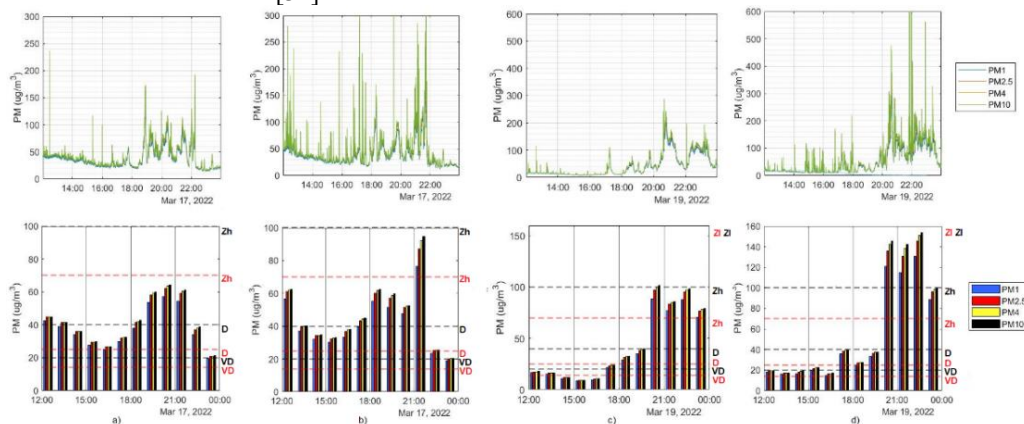
Z Obr. 4 je zrejme, že hoci meracia stanica S2 (Obr. 4b) namerala určité špičky, ktoré trvali krátku dobu (vo väčšine prípadov s nábehom a poklesom niekoľko minút), tieto špičky nie sú ani zďaleka také vysoké a ani také časté, ako to je u meracej stanice S1 (Obr. 4a). Najväčšia špička koncentrácie PM10, ktorú zariadenie S2 nameralo, malo cca $900 \mu\text{g}/\text{cm}^3$, zatiaľ, čo zariadenie S1 zosnívalo niekoľko špičiek, ktoré presahujú túto hodnotu. Najväčšia špička nameraná S1 presahuje hodnotu $2600 \mu\text{g}/\text{cm}^3$. Dôležité je však podotknúť, že hodnota hmotnostnej koncentrácie PM10 mimo trvania špičiek je porovnateľná u oboch meracích zariadení, čo môžeme vidieť aj vo vybraných intervaloch merania na Obr. 5.



Obr. 5 Porovnanie nameraných hodnôt senzorov SPS30 (S1 a S2). Intervaly a) a f) neobsahujú vysoké špičky a priebehy majú podobný charakter s tým, že viac znečistené ovzdušie je na a). Intervaly b) – e) nazorujú špičky namerané zariadením S1, spôsobené blízkym umiestnením od zdroja pevných častíc. Zariadenie S2 tieto špičky nezosiťvalo, keďže vzdialenosť od zdroja pevných častíc bola väčšia.

Pri pozorovaní vyššie uvedených intervalov sa nám odhaľuje ešte ďalšia otázka: Do akej miery ovplyvňujú krátko trvajúce, avšak časté vysoké špičky v hmotnostnej koncentrácii pevných častíc

celkovú kvalitu ovzdušia? Pre jasnejšie zhodnotenie kvality ovzdušia boli vypočítané hodinové priemery koncentrácií PM. Obr. 6 obsahuje na hornom rade obrázkov namerané priebehy PM1, PM2,5, PM4 a PM10 pre vybrané intervaly. Dolný rad obrázkov tvoria stĺpcové grafy obsahujúce priemerné hodnoty. Čiarkovaná čiara sú zaznačené limitné hodnoty. Červené čiarkované čiary sa vzťahujú na PM2,5 a čierne na PM10, pričom VD = veľmi dobrá, D = dobrá, Zh = zhoršená, Zl = zlá kvalita ovzdušia [32].



Obr. 6 Porovnanie kvality ovzdušia na balkóne a v záhrade.

17. 3. 2022, 12:00 – 24:00 na: a) záhrade, b) balkóne. Tento interval je rozšírením intervalu na Obr. 5c).
19. 3. 2022, 12:00 – 24:00 na: c) záhrade, d) balkóne. Tento interval je rozšírením intervalu na Obr. 5e).

Z porovnania Obr. 6a) a Obr. 6b) vyplýva, že vysoké špičky (samotná špička trvá od 5 do 60s, ale významný nárast a pokles nameraných PM na Obr. 4a trvá 2 až 8 min.) ovplyvňujú aj celkovú kvalitu ovzdušia. Najväčší rozdiel pre tento interval je v 21. hodine. Podobné výsledky pozorujeme aj pri Obr. 6c) a Obr. 6d), kde je však zvýšená hmotnostná koncentrácia PM10 ešte výraznejšia, nakoľko na Obr. 6d) hmotnostná koncentrácia PM10 prekračuje hornú hranicu zhoršenej (Zh) kvality ovzdušia. Stav zlej (Zl) kvality ovzdušia trvá až tri hodiny, konkrétne od 20. do 22. hodiny vrátane. Počas 23. hodiny hmotnostná koncentrácia PM10 opäť klesá, no stále sa približuje k hornej hranici zhoršenej (Zh) kvality ovzdušia. Dokonca aj v odpoľudňajších hodinách (14. – 17. hod.), kedy špičky dosahujú „len“ $100 \mu\text{g}/\text{cm}^3$, sú tieto špičky postačujúce na prekročenie hornej hranice pre veľmi dobrú (VD) kvalitu ovzdušia. Z týchto výsledkov je možné usúdiť, že vzdialenosť od zdroja PM je významným faktorom pre kvalitu ovzdušia. Aj rozdiel vzdialeností 35 m má na ňu značný vplyv. Takéto podmienky spôsobené vykurovaním sú nebezpečné najmä pre seniorov a deti žijúcich na vidieku a pre ich zdravie.

IV. ZÁVER

Z minulých skúseností so senzormi (napr. senzormi teploty a tlaku BMP280 či senzormi oxidu uhličitého MQ-135) sme zistili, že nie všetky senzory majú dodržané výrobcom stanovené tolerancie, čo bol dôvod, prečo boli vykonané porovnávacie merania so senzormi pevných častíc SPS30. Z porovnávacích meraní senzorov SPS30 vyplynulo, že senzory spadajú do tolerancie uvedenej výrobcom, a teda je možné pokladať výsledky namerané týmito senzormi za spoľahlivé.

Meranie na dvoch rôznych miestach na vidieku a reálne horšie dýchateľný vzduch v blízkosti rodinného domu ako na záhrade, potvrdili veľký vplyv zmeny veľkosti koncentrácie PM (časové zmeny PM ale aj hodinové priemery vo forme stĺpcových grafov – Obr. 6) v závislosti na blízkosti komínov rodinných domov. Konkrétne, merania vykonané na balkóne zaznamenali omnoho vyššie špičky v hmotnostnej koncentracii pevných častíc než merania vykonané v záhrade, ktoré boli viacej vzdialené od komínov rodinných domov. Výpočet hodinových priemerov hmotnostnej koncentrácie pevných častíc umožňujú klasifikáciu kvality ovzdušia podľa limitných hodnôt stanovených [32]. Prítomnosť vysokých špičiek v hmotnostnej koncentracii pevných častíc prispela k zhoršenej až zlej kvalite ovzdušia na Obr. 6.

V súčasnosti, keď je nosnou témou zmena energetickej politiky, ochrana životného prostredia či stúpajúce náklady na energie sa bude drevo (napriek rastúcim cenám aj tejto komodity), javiť ako najpriateľnejší, najlacnejší a najdostupnejší zdroj tepla. Je to aj preto, lebo drevo ako surovina je produkované v mnohých oblastiach regionálne s minimálnymi požiadavkami na transport k odberateľovi. Vykurovanie drevom vo väčšej miere ako to bolo pred rokom 2022 (vďaka stúpajúcim cenám plynu a elektrickej energie) už má za následok iba obmedzené použitie výhod rekuperácie vzduchu v rodinných domoch, či väčšie znečistenie ovzdušia na dedinách v porovnaní napr. s mestom Košice, kde hlavným zdrojom znečistenia ovzdušia je dlhodobá cestná premávka a US Steel. Aj preto je potrebné sa zaoberať meraním PM, zberom údajov a vyhodnocovaním a zistiť možnú koreláciu medzi znečistením ovzdušia a poveternostnými

podmienkami, čo by pomohlo pri predikcii kvality ovzdušia vo veľkých mestách aj na vidieku. To však bude predmetom našich budúcich aktivít.

POĎAKOVANIE

Táto publikácia bola podporená zo zdrojov Grant FEI-2022-82.

POUŽITÁ LITERATÚRA

- [1] J. Fraden, *Handbook of Modern Densors: Physics, Designs, and Applications*, 5. ed. Cham: Springer, 2016. ISBN: 978-3-319-19302-1.
- [2] BOSCH, "Digital Pressure Sensor", BMP280 Datasheet, Máj 2015, Spôsob prístupu: <https://cdn-shop.adafruit.com/datasheets/BST-BMP280-DS001-11.pdf>
- [3] B. G. Lipták, *Instrument Engineers' Handbook: Process Measurement and Analysis*, 4. ed., Boca Raton, FL: CRC Press, 2003. ISBN: 0-8493-1083-0.
- [4] BOSCH, "Digital Pressure Sensor", BMP384 Datasheet, Jan. 2018, Spôsob prístupu: <https://cdn.sos.sk/productdata/f0/3e/9e/9b/5bb9/bmp384.pdf>
- [5] BOSCH, "Digital Pressure Sensor", BMP388 Datasheet, Nov. 2020, Spôsob prístupu: <https://www.mouser.com/pdfdocs/BST-BMP388-DS001-01.pdf>
- [6] "Gas sensor", MQ135 Datasheet, Spôsob prístupu: <https://www.olimex.com/Products/Components/Sensors/Gas/SNS-MQ135/resources/SNS-MQ135.pdf>
- [7] M. Guzan, "Meranie teploty s SHT31 Smart GADGET", SOS electronic Roč. 1, č. 2, 2019, s. 8-11. - ISSN 2644-5654 Spôsob prístupu: https://www.sos.sk/novinky/pdf/2379_sos-news-2019-02-sk-web.pdf.
- [8] M. Guzan, A. Fehér, R. Mitro, "Meranie teploty a vlhkosti aj senzormi Sensirion v praxi", SOS electronic Roč. 2, 2020, s. 27-27. - ISSN 2644-5654 Spôsob prístupu: <https://www.sos.sk/articles/sos-supplier-of-solution/meranie-teploty-a-vlhkosti-aj-senzormi-sensirion-2499>.
- [9] R. Mitro: *Spresnené meranie teploty*, Diplomová práca, Vedúci: doc. Ing. Milan Guzan, PhD., Konzultant: Ing. Simona Kirešová, FEI TUKE, 2022, s.75.
- [10] H. Kwoon, M. Ryu, C. Carlsten, "Ultrafine particles: unique physiochemical properties relevant to health and disease", *Experimental and Molecular Medicine*, Roč. 52, s. 318-328, Mar. 2020.
- [11] S. Y. Kyung, S. H. Jeong, "Particulate-Matter Related Respiratory Diseases", *Tuberculosis and Respiratory Diseases*, Roč. 83, č. 2, s. 116-121, Mar. 2020.
- [12] E. J. Jo et al., "Effects of particulate matter on respiratory disease and the impact of meteorological factors in Busan, Korea", *Respiratory Medicine*, Roč. 124, s. 79-87, Feb. 2017.
- [13] T. Götschi, J. Heinrich, J. Sunyer, N. Künzli, "Long-term effects of ambient air pollution on lung function: A review", *Epidemiology*, Roč. 19, č. 5, s. 690-701, Sep. 2008.
- [14] R. D. Brook, "Cardiovascular effects of air pollution", *Clin Sci (Lond)*, Roč. 115, č. 6, s. 175-187, Aug. 2008.
- [15] C. Mühlfeld, B. Rothen-Rutishauser, F. Blank, D. Vanhecke, M. Ochs, P. Gehr, "Interactions of nanoparticles with pulmonary structures and cellular responses", *The American Journal of Physiology-Lung Cellular and Molecular Physiology*, Roč. 294, č. 5, L817-L829, Máj 2008.
- [16] M. Kampa, E. Castanas, "Human health effects of air pollution", *Environmental Pollution*, Roč. 151, č. 2, s. 362-367, Jan. 2008.
- [17] D. E. Schraufnagel, "The health effects of ultrafine particles", *Experimental & Molecular Medicine*, vol. 52, pp. 311-317, Mar. 2020.
- [18] S. Y. Kyung, S. H. Jeong, "Particulate-Matter Related Respiratory Diseases", *Tuberculosis and Respiratory Diseases*, Roč. 83, č. 2, s. 116-121, Mar. 2020.
- [19] E. J. Jo et al., "Effects of particulate matter on respiratory disease and the impact of meteorological factors in Busan, Korea", *Respiratory Medicine*, Roč. 124, s. 79-87, Feb. 2017.
- [20] T. Götschi, J. Heinrich, J. Sunyer, N. Künzli, "Long-term effects of ambient air pollution on lung function: A review", *Epidemiology*, Roč. 19, č. 5, s. 690-701, Sep. 2008.
- [21] R. D. Brook, "Cardiovascular effects of air pollution", *Clin Sci (Lond)*, Roč. 115, č. 6, s. 175-187, Aug. 2008.
- [22] R. M. Harrison et al., *Airborne particulate matter: sources, atmospheric processes and health*, Royal Society of Chemistry, UK, 2016, ISBN: 978-1-78262-491-2.
- [23] Z. Jonáček, M. Zemko, J. Szemesová, L. Zetochová, "Quality Improvements of Air Emission Accounts and Extension of Provided Time-Series Focusing on Households Heating", *Meteorological Journal*, Roč. 23, č. 1, s. 41-46, Jul. 2020.
- [24] A. Suleiman, M. R. Tight, A. D. Quinn, "Assessment and prediction of the impact of road transport on ambient concentrations of particulate matter PM10", *Transportation Research Part D: Transport and Environment*, Roč. 49, s. 301-312, Dec. 2016.
- [25] D. Štefánik, J. Matejovičová, "Mean Exposure to PM2.5 Particles in Slovakia in 2017 and Number of Premature Deaths", *Meteorological Journal*, Roč. 23, č. 1, s. 31-40, Jul. 2020. (In Slovak).
- [26] R. Ravi Krishna., S. M. Shiva Nagendra, M. Saraswati, Diya, *Urban Air Quality Monitoring, Modelling and Human Exposure Assessment*, Springer, Singapore, 2020, ISBN: 978-981-15-5510-7.
- [27] S. S. Amaral, J. A. De Carvalho Jr., M. A. Martins Costa, C. Pinheiro, "An Overview of Particulate Matter Measurement Instruments", *Atmosphere*, Roč. 6, č. 9, s. 1327-1345, Sep. 2015.
- [28] S. D. Lowther, K. C. Jones, X. Wang, J. D. Whyatt, O. Wild, D. Booker, "Particulate Matter Measurement Indoors: A Review of Metrics, Sensors, Needs, and Applications", *Environmental Science & Technology*, Roč. 53, č. 20, pp. 11644-11656, Oct. 2019.
- [29] B. Denker, E. Shklovsky, Ed., *Handbook of solid-state lasers: Materials, systems and applications*, Cambridge, England, Woodhead Publishing, 2013, ISBN: 978-0-85709-272-4.
- [30] Sensirion, "Automotive Grade Humidity and Temperature Sensor", SHT3xA-DIS datasheet Dec. 2019. Spôsob prístupu: <https://cdn.sos.sk/productdata/98/89/92718144/sps30-2.pdf>
- [31] Z. Chafe et al., "Residential heating with wood and coal: health impacts and policy options in Europe and North America", World Health Organization, 2015. Spôsob prístupu: <https://www.euro.who.int/en/publications/abstracts/residential-heating-with-wood-and-coal-health-impacts-and-policy-options-in-europe-and-north-america>
- [32] Slovenský hydrometeorologický ústav. (2022, Feb 4). *Hodinové koncentrácie znečisťujúcich látok* [Online]. Spôsob prístupu: https://shmu.sk/sk/?page=1&id=oko_imis.

Digitálna vodotlač zabezpečujúca obrazové dáta pred útokmi vyrezaním

¹Samuel ANDREJČÍK, ²Luboš OVSENIK, ³Jakub ORAVEC,
⁴Norbert ZDRAVECKÝ, ⁵Maroš LAPČÁK

^{1,2,3,4,5} Katedra elektroniky a multimediálnych telekomunikácií, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

¹samuel.andrejck@tuke.sk, ²lubos.ovsenik@tuke.sk, ³jakub.oravec@tuke.sk,
⁴norbert.zdravecky@tuke.sk, ⁵maros.lapcak@tuke.sk

Abstrakt — Tento príspevok rozoberá digitálnu reprezentáciu obrazov, históriu a princípy techník zabezpečenia obrazových dát, konkrétne digitálnu vodotlač a steganografiu. Taktiež bola navrhnutá implementácia krehkej digitálnej vodotlače a následne vyhodnotená údajmi z útoku vyrezaním menšej a väčšej časti obrazu, ktorá na označenie dát vodoznakom využíva modifikácie najmenej významnej bitovej roviny obrázka.

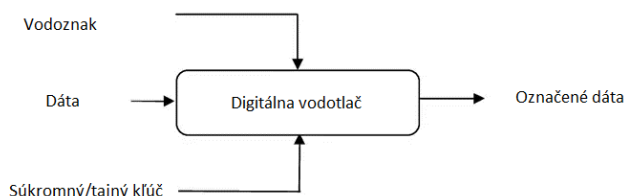
Kľúčové slová — digitálna vodotlač, LSB, steganografia, vodoznak, útok

I. ÚVOD

Internet sa stal denne vyhľadávanou službou a v súčasnosti si drvivá väčšina ľudí už nevie predstaviť život bez neho. Jeho využívanie ponúka nespočetné množstvo vymožeností, od kontaktu ľudí, ktorí sú pokojne aj na opačných stranách sveta až po vyhľadávanie dôležitých informácií a obsahu, za ktorým nemusíme nutne cestovať do knižnice. Na internete nájdeme široké spektrum obsahu, nie však všetok je nám k úžitku. Určite skupiny ľudí totižto vedia jednotlivé služby internetu upraviť tak, aby nám uškodili. Takouto na prvý pohľad neodhaliteľnou úpravou môžu pomocou takýchto služieb zneužiť internet na krádeže či modifikácie údajov a informácií, ktoré sa po sieti odosielaajú. Tento problém však rieši digitálna vodotlač, ktorá bola vytvorená za účelom zabezpečenia zachovania autenticity a bezpečnosti dát, ako aj autorských práv digitálnych multimédií. Motiváciou k riešeniu tohto problému nám bola skutočnosť, že sa na internete zdieľa obrovské množstvo dát, ktoré sú reprezentované digitálnymi multimédiami. Bezpečnosť takto preposielaných dát je častokrát veľmi nízka a to otvára dvere útočníkom, ktorí môžu dáta odcudziť, resp. ich modifikovať. Riešenie problému by mohlo predstavovať navrhnutie implementácie krehkej vodotlače, ktorá dokáže vystopovať manipuláciu s obrazom a rovnako tak dokáže zabezpečiť autorské práva. Takáto technika využíva modifikácie najmenej významnej bitovej roviny obrázka. Z historického hľadiska bola digitálna vodotlač vyvinutá už v staroveku z techniky steganografie. Už v tej dobe bola myšlienka utajiť informácie v komunikácii veľmi častá a jej realizovanie veľmi dôležité. Práve slovo steganografia, ktoré sa dnes ešte stále používa, pochádza z gréckeho jazyka a znamená utajenú komunikáciu [1]. Samotnú steganografiu možno opísať ako možnosť ukrytia dát v obsahu multimédií, čo znamená, že namiesto toho aby sa dáta zašifrovali sa len vložia do väčšieho objemu dát. Takýto proces nevyžaduje tajný prenos a prenášané dáta takto slúžia ako nosič tajnej správy. Cieľom je ukryť dáta tak, aby príjemca na konci prenosu bol schopný extrahovať obsah dát, nakoľko vie, čo má v prenášaných dátach hľadať. Z histórie poznáme nespočetné množstvo typov takto prenášaných dát, ktoré boli ukryté v kresbách, upravených slovách, prípadne vpísané pod vlasy človeka, či kožušinu zvieratá.

II. PRÁCA S DIGITÁLNOU VODOTLAČOU

Digitálna vodotlač vloží informácie do digitálnych údajov tajným a nenápadným spôsobom. Vložené informácie sa označujú ako digitálny vodoznak [2]. Digitálny vodoznak musí byť robustný voči skresleniu, ktoré významne neznehodnocuje vnímanie nosného signálu prijímačom [3]. Základnou vlastnosťou vodotlače je to, aby bol vložený vodoznak v prenášaných dátach nenápadný a bezpečný. Použitím šifrovacieho kľúča je možné vodoznak do obrazových dát pridať, ale aj odobrať. Pridaním vodoznaku do dát vzniká digitálna vodotlač (Obr. 1).

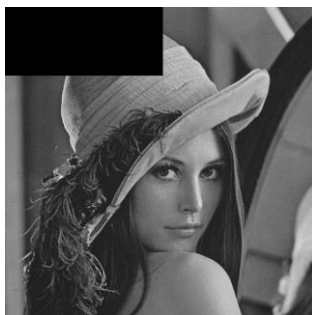


Obr. 1 Schéma znázorňujúca vytváranie digitálnej vodotlače

Existujú rôzne možné úmyselné alebo neúmyselné útoky, ktoré sa pravdepodobne dotknú objektu označeného vodoznakom. Dostupnosť širokej škály softvérov na spracovanie obrazu umožnila útoky na robustnosť systémov vodotlače. Cieľom týchto útokov je zabrániť vodoznakovi pracovať spôsobom preňho typickým [4].

III. ÚTOK VYREZANÍM MENŠEJ ČASTI OBRÁZKA

Na navrhnutý systém ochrany obrazových dát bol realizovaný útok vyrezaním menšej časti obrázka. Pri vykonaní útoku bol použitý obrázok v odtieňoch šedej *lena512g.bmp* s rozlíšením 512x512 pixelov obrazových prvkov, program *Irfanview*, pomocou ktorého bola menšia časť obrázka vyrezaná a logo, reprezentujúce vodoznak, v odtieňoch šedej s rozlíšením 64x64 pixelov obrazových prvkov. Výsledok tohto útoku je odprezentovaný aj na nižšie priložených obrázkoch. Výsledky dosiahnuté pri realizácii útoku na ďalšie obrázky sú odprezentované v priloženej Tabuľke 1. Je dôležité si uvedomiť, že pri realizovaní útoku vyrezaním časti obrázka nie je jedno, akú časť obrázka vyrežeme. Pri vyrezaní strednej časti obrázka môžeme pri jednom obrázku stratiť dôležitú informáciu, kým pri inom obrázku neprídeme o zásadné dáta. Pri realizácii útoku vyrezaním bol z obrázka vybraný ľavý horný roh (Obr. 2).

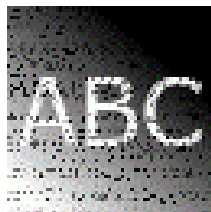


Obr. 2 Testovaný obrázok po zrealizovanom útoku vyrezaním menšej časti obrázka

Ak by implementácia krehkej vodotlače nepoužila posun obrazových prvkov pred tým, ako sa vložil vodoznak, tak by väčšina z neho bola odstránená po realizácii útoku. Na základe Obr. 3 a Obr. 4 je možné posúdiť efektívnosť posunu obrazových prvkov, keďže nebola porušená celistvosť vodoznaku.



Obr. 3 Extrakcia vodoznaku pred zrealizovaným útokom



Obr. 4 Extrakcia vodoznaku po zrealizovanom útoku

Porovnaním extrakcií vodoznakov z obrázka, na ktorý bol vykonaný útok vyrezaním malej časti obrázka je možné tvrdiť, že navrhnutá krehká vodotlač je dostatočne odolná voči takémuto

typu útoku. Aj napriek tomu, že niektoré obrazové prvky stratili svoju hodnotu, vďaka posunutiu obrazových prvkov sa táto strata rozprestrela na celý obrázok a práve tým sa stala zanedbateľnou (Obr. 4), nakoľko nespôsobila stratu celého vodoznaku. Degradáciu vodoznaku je možné pozorovať prostredníctvom meraného parametra – normovaného korelačného koeficientu. Zo získaných údajov, ktoré sú odprezentované v Tabuľke 1 sa dá potvrdiť efektívna funkčnosť implementovanej krehkej vodotlače.

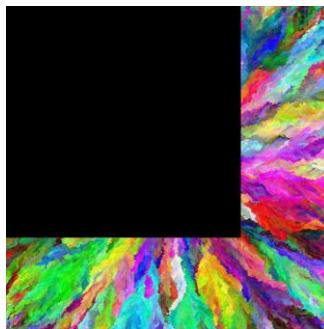
Tabuľka 1
Prezentácia údajov získaných útokom vyrezania menšej časti obrazu

Obrázok	Rozmery vyrezanej časti (px)	Normovaný korelačný koeficient	Degradácia vodoznaku
colorful512c.bmp	112x260	0,8942	Jemná
lena512g.bmp	260x112	0,9235	Minimálna
airplane512c.bmp	258x110	0,9289	Minimálna
couple512g.bmp	110x258	0,9272	Minimálna
peppers512c.bmp	360x63	0,9533	Minimálna
boat512g.bmp	63x360	0,9569	Minimálna

Na základe údajov zapísaných v Tabuľke 1 je možné tvrdiť, že vyrezaním malej časti obrazu nie je možné vodoznak poškodiť natoľko, aby stratil svoj zmysel. Priemerná hodnota normovaného korelačného systému bola v približnom priemere 0,93, čo potvrdzuje toto tvrdenie. Útoky degradovali vodoznak len minimálne a teda výsledkom je, že aj pri realizácii takéhoto útoku je stále možné overiť autorské práva, alebo zistiť či niekto s obrázkom manipuloval.

IV. ÚTOK VYREZANÍM VÄČŠEJ ČASTI OBRÁZKA

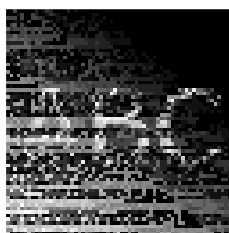
Druhé testovanie navrhovaného riešenia zabezpečenia obrazových dát prostredníctvom krehkej vodotlače bolo realizované útokom vyrezania väčšej časti obrázka (Obr. 5). Pri realizácii útoku boli opäť použité vyššie merané obrazy a vodoznak, ktorý bol prezentovaný logom v odtieňoch šedej s rozlíšením 64x64 pixelov. Výsledok tohto útoku je odprezentovaný graficky na nižšie priložených obrázkoch, pričom prezentované výsledky boli realizované na obrázku *colorful512c.bmp*.



Obr. 5 Testovaný obrázok po zrealizovanom útokom vyrezaním väčšej časti obrázka



Obr. 6 Extrakcia vodoznaku pred zrealizovaným útokom



Obr. 7 Extrakcia vodoznaku po zrealizovanom útokom

Pri porovnaní extrakcií z Obr. 6 a Obr. 7 je možné usúdiť, že krehká vodotlač už nie je dostatočne odolná voči útoku vyrezaním väčšej časti obrázka. Vodoznak je kvôli útoky v najmenej významnej bitovej rovine natoľko poškodený, že je ťažšie ho spozorovať. Z tohto dôvodu je možné tvrdiť, že vyrezanie väčšej časti obrazu poškodí vodoznak do takej miery, že porovnanie s pôvodným vodoznakom môže byť náročné. Výsledky dosiahnuté pri realizácii útoku na ďalšie obrázky sú odprezentované v priloženej Tabuľke 2. Nižšie odprezentované výsledky potvrdzujú, že pri tomto type útoku je nameraná priemerná hodnota normovaného korelačného koeficientu približne na úrovni 0,57 a tým je možné prehlásiť, že navrhnutý systém nie je dostatočne robustný na to, aby vydržal tak masívny útok. Podobne, ako pri predchádzajúcom útoky sú výsledky popísané na 6 rôznych obrazoch a útoky na nich boli realizované s rôznymi rozmermi vyrezanej časti obrázka. Keďže oproti predošlému pokusu boli rozmery výrezov 5 násobné, priamo úmerne sa znižovala aj kvalita a celistvosť extrahovaných vodoznakov.

Tabuľka 2
Prezentácia údajov získaných útokom vyrezania väčšej časti obrazu

Obrázok	Rozmery vyrezanej časti (px)	Normovaný korelačný koeficient	Degradácia vodoznaku
colorful512c.bmp	369x365	0,6452	Stredná
lena512g.bmp	260x112	0,6481	Stredná
airplane512c.bmp	258x110	0,5632	Stredná
couple512g.bmp	110x258	0,6234	Stredná
peppers512c.bmp	360x63	0,4922	Horšia
boat512g.bmp	63x360	0,4851	Horšia

V. ZÁVER

Implementáciu krehkej digitálnej vodotlače sa nám podarilo úspešne otestovať. Pri vyrezaní menšej časti obrázka sa hodnoty normovaného korelačného koeficientu, ktorý bol využitý na porovnanie s pôvodným vodoznakom, takmer v každom prípade skoro rovnali 1, čo predstavuje úplnú zhodu. To znamená, že vyrezanie malej časti obrázka, bez ohľadu na miesto výrezu, nespôsobí veľké straty. Avšak, pri vyrezaní väčšej časti obrázka sú hodnoty normovaného korelačného koeficientu príliš nízke. Taktiež pri vizuálnej kontrole vodoznakov bolo vidieť, že kvalita a čitateľnosť vodoznakov bola mizivá. Ak by nebol využitý posun obrazových prvkov, tak by väčšina útokov vyrezaním dopadla oveľa horšie. Keďže posun obrazových prvkov zabezpečí rozmiestnenie vodoznaku po celej najmenej významnej bitovej rovine, tak pri každom výreze sa odstráni len nejaká časť vodoznaku, nie celý. Možným vylepšením implementácie by mohlo byť označenie celej najmenej významnej bitovej roviny, čím by sa vodoznak vložil niekoľkokrát po sebe, nie len raz. To by zaručilo väčšiu robustnosť voči útokom vyrezaním časti obrázka a zvýšilo by to odolnosť voči iným útokom.

POĎAKOVANIE

Táto práca bola podporená výskumným projektom FEI-2022-84 "Techniky spracovania dát vo vysokorýchlostných prenosových systémoch".

LITERATÚRA

- [1] S. Kolkus et al., „Human Skin Detection Using RGB, HSV and YCbCr Color Models“, 2016. <https://www.atlantispress.com/article/25871632.pdf>.
- [2] B. Furht, et al. „Multimedia Encryption and Watermarking“, 2005. 327 s. ISBN: 978-0-387-24425-9.
- [3] Ch. Busch et al., „Digital Watermarking: From Concepts To Real-Time Video Applications“, 1999. ISSN: 1558-1756.
- [4] P. Singh, R. Chadha, „A Survey of Digital Watermarking Techniques and its Applications“, 2013. https://www.academia.edu/7375437/A_Survey_of_Digital_Watermarking_Techniques_and_its_Applications

Web-based Solution for Management of Records

¹Jakub FRANKOVIČ, ²Ondrej KAINZ, ³Miroslav MICHALKO,
⁴František JAKAB, ⁵Ivana NOVÁKOVÁ

^{1,2,3,4,5}Department of Computers and Informatics, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Slovak Republic

¹jakub.frankovic@student.tuke.sk, ²ondrej.kainz@tuke.sk,
³miroslav.michalko@tuke.sk, ⁴frantisek.jakab@tuke.sk, ⁵ivana.novakova@tuke.sk

Abstract — The paper describes development of a small content management system, which comes with a public interface for users and will be deployed in the university environment. Prior to the implementation of this system, the analysis of the web services, CMS systems and also PHP frameworks was carried out. Based on the analysis, a system that includes an interface for administrators and for basic users was developed. The implementation of the server part was built on the Laravel framework, which communicates with the MySQL relational database. On the client side, JavaScript has mostly been used in combination with the JQuery library and other, smaller libraries. The design was created using the CSS framework Bootstrap and SASS technology. The created solution was tested during the entire development and after the end of the development by a sequence of recommended tests. The output product is a fully functional web-based system for managing and displaying of the records.

Keywords — JavaScript, Laravel, MySQL, PHP, web application.

I. INTRODUCTION

Content management systems are very popular in the field of web technologies today. In the past, managing data without these systems took a long time. The developed systems shortened this process to a few clicks of the responsible person - the administrator. The solutions that preceded the actual implementation of such systems have been described by Nakwaski and Zabierowski [1]. At the time of the growth of the Internet, websites were mostly static, and the content was only simple contact information. This means that they were not linked to any database and did not download additional content - so their creation was relatively simple and required only knowledge of HTML or CSS. However, over time, business owners and virtually different website owners have come to conclusion that for a website to be high quality, it must be up to date. This idea was the beginning of the dynamization of the web.

Managing dynamic websites was not easy for their programmers and administrators - more programming skills were required, and the code also became more unreadable. This led to the idea of creating the already mentioned record management systems, which would rapidly speed up the development of websites, but also their administration and continuous updating. describes how the eye tracking data can be visualized and thus used for various purposes of attention measuring and more.

The role of programmers has also changed - they will no longer form all the parts of a website but will instead develop tools that do it automatically. The development of such tools at that time was also helped by the trend of programming languages, which became more and more advanced, and all this resulted in the completion of this idea. Currently, these systems are enjoying immense popularity. This is also confirmed by statistics, which say that out of the most monitored 1,000,000 pages of the Internet, 761,802 use a CMS system. No wonder - these systems offer a simple and fast way to create an advanced website that is easy to administer. Today, Wordpress is by far the most widely used CMS. Of the 1,000,000 pages that run on CMS, about 33% are Wordpress. In addition, this system is free, which certainly adds to its dominance in today's market of record management systems.

II. WEB SERVICES

A web service is a solution for how applications can communicate with each other and exchange information with each other over the Internet [2]. In other words, based on the definitions from

sources [3], [4] and [5], they are software systems or applications that interact with client applications and other services using open standards for the purpose of data exchange. They combine the best aspects of web and component programming, regardless of the programming language, operating system, and component model in which they were implemented.

A. Importance and use of web services

A web service that acts as a server provides the client with various functionality, such as computational and data processing operations, in the form of encapsulated services [5]. With the help of web services, this functionality becomes available through promotional mechanisms, which are often in the form of computer-readable documents and describe all the operations that this service provides. These services do not have to be deployed only on the World Wide Web (WWW) network structure, but on any network interface structure.

One of the basic features of web services is the ability of computers to work together with the provision of services (interoperability) and further scalability [5]. These features are achieved using web technologies such as Extensible Markup Language (XML) and Hypertext Transfer Protocol (HTTP). The use of these technologies greatly reduces dependence on proprietary networks and platforms. Thanks to this, software applications can very easily connect to such services. Just as we can connect an application to a web service, we can connect another web service to it, creating more complex prototypes of web services, but offering more functionality. An example is the web service of an online store - it can integrate other services for payments (various payment gateways), transport services and storage.

The correct use of web services depends in part on the application area. These are quite often provided as distributed systems. This means that the provider and the client are located in a completely different physical location and communicate over the network - so it is necessary to point out that the service provided is slower and less reliable than the service provided in the local system - in other words, we need to take network unpredictability into account. These negative features must therefore be considered when implementing web services in our application.

B. Standards

ML, SOAP, WSDL and UDDI [6] are essential elements in the deployment of web services-based SOA infrastructures. XML is the standard language for data representation. SOAP is a connection between users and providers. WSDL describes web services and UDDI is used to register and search web services.

XML

XML was introduced to represent web services data. It is accepted as a standard for exchanging data on the web, enabling their structuring. It is a language for so-called semi-structured data and its idea is to solve the problem of data integration - it allows flexible coding and display of data using metadata, which describe the structure of the data stored in XML [7].

For all messaging systems, the choice of the medium that transmits this information is an important decision. A common understanding of what exactly a report is needed. In web services, according to [8], a message is an XML document that is described and defined by an XML infoset. Infoset is a theoretical data model that is compatible with XML 1.0 and is the basis of all modern specifications of this language (XML Schema, XML Query, etc.). The XML document is composed of so-called named containers and their associated data. Typically, these containers are represented as declarations, elements, and attributes. The declaration specifies the XML version by which the document is defined. The technical term for a text unit in XML is an element. Element containers can be defined to contain data, other elements, data, and other elements together, or need not contain anything at all. An XML document is also known as an instance of an XML document. This instance represents a possible set of data for a specific mark-up language.

SOAP

SOAP is an XML-based communication protocol for exchanging messages between computers regardless of their operating systems or programming environment. It is a standard messaging protocol used by web services, characterizes Papazoglou in [9]. The primary function of this protocol is inter-application communication. SOAP assumes the use of XML as a schema for encoding requests and responses using HTTP as a medium. The SOAP method is simply an HTTP request and response that complies with SOAP encryption rules. A SOAP endpoint is an HTTP-based URL that identifies the destination for invoking the method. SOAP was originally an object-oriented access protocol but does not require any object-oriented access. Rather, it

defines a model for using simple request and response messages written in XML as the basic protocol for electronic communication.

SOAP can be called a wire protocol because it determines how messages associated with services are structured over the Internet. It can also be defined as a lightweight protocol designed to exchange structured and type information between different systems in an environment such as Internet, or some LAN, in order to invoke the remote method. It can send and receive HTTP packets and can process XML messages. In contrast, we can mention the protocols of the architecture of distributed objects, such as Internet Inter-ORB Protocol, where a suitable run-time environment must be installed for it to work properly and users must also configure their systems to meet the requirements of these systems [9]. Lightweight protocols that are not considered by the environment, programming language, and operating system are better at this. According to the author of the book [9], many people cannot distinguish between wire protocol and transport protocol. The difference between them is that the wire protocol specifies the form in which data is exchanged between two different applications. Transport protocol is a method by which data is exchanged between systems. He is also responsible for getting the data from the sender to the destination.

III. CMS SYSTEMS

Today, the market for CMS systems is very diverse. The availability of free solutions as well as cheap web hosting is influenced by the fact that not only companies, but also ordinary people can create a website relatively easily. The author [1] further states that these systems have experienced a great increase in popularity since their inception. For example, an article currents up to December 4, 2020, points out that more than 35.6% of all sites on the Internet are created using the WordPress system.

Web content management is a set of processes that take care of digital information. We can also call this information more specifically as digital content. It can take the form of text, multimedia files, such as images, videos, or audio.

We don't have to understand content management systems only in the context of websites. Based on the book [10], the authors [1] say that the main difference of these systems, or their groups, is the functionality they provide and based on that they have a different structure. There are systems for managing the content of websites, documents, transactions, publications, education, enterprise content or integration.

A. CMS Architecture

A content management system is a system that allows authorized people to edit, update, and delete digital content without being highly technologically experienced. The author [1] says that from a security point of view, not every such user can make any adjustments they can think of. Users are limited by rights, so if they want to delete some sensitive data, for example, they will not succeed because they will not have the right to do so.

Such systems are in most cases based on a database that maintains both the public information that is displayed to site visitors and the configuration data for scripting languages that run on the server side. Based on a wide range of web hosting providers, and thus different pricing options, these systems have tried to use the most affordable solutions possible, continues CMS Zabierowski in [1]. The most accessible came from a combination of PHP scripting language and MySQL database.

Content management systems use predefined templates to display data. The template consists of HTML elements, CSS styles and JavaScript scripting language. We can create these templates either ourselves or use templates that are publicly available. The content is separate from the template, it is in the database. The server part of the web application retrieves this data from the database and displays it in predetermined places on the web. The authors [1] further talk about the biggest advantage of separating content from a template. This is that if we do not like the design, we can easily replace it, i.e., modify it without the need for changes in the content page, for example by changing the parameters of the table in the database. According to [1], content management systems are usually divided into a presentation part and an administrative part. We can imagine the presentation part as any website we visit. This section is accessible to everyone and offers information that is supervised and cared for by the website administrator. In content management systems, the administrator uses the latter administrative part to manage this publicly available information. It is also available via the website for all users who can authenticate and then have access to various modifications of the website. This is not a requirement, but content management systems often include functionality that allows individual administrators to still assign rights. This means that, for example, a news site administrator can edit articles but does not have the right to delete them.

B. Technologies

Technologies are constantly evolving, so we need to keep up with them and use modern technologies for solutions. Based on the architecture of the developed solution, it is important to choose a scripting language and database in which the data will be stored. The author of the article [11] writes that CMS uses scripting, resp. programming language for obtaining information stored in the database, formatting and subsequent submission to the browser. In this case, different programming languages can be used. A well-known and proven combination is a pair of PHP and MySQL. Although the global dominance of PHP is declining, it is still the preferred language for creating websites. Especially in combination with the popular and constantly evolving and updated PHP frameworks, it is a competitive language.

IV. PHP Frameworks

If we wanted to develop the web using ordinary and pure PHP, it is very likely that we would start mixing software logic, working with the database and a presentation layer in the form of HTML tags. As a result, maintenance and scalability of the application are close to impossible [12]. PHP frameworks have been developed to address this issue. These help developers build web applications faster, easier, and using a specific model that uses the framework. Preset APIs, libraries, and tools may also be available to help developers produce better quality code. In most cases, PHP frameworks are based on the Model View Controller (MVC) design model. This model is an efficient and proven way to generate modular and organized applications. MVC as a design model typically divides the application into separate layers, which can be analysed and implemented often separately, the author writes [13]. He adds that by splitting models and views, MVC helps reduce the complexity of architectural design and increase flexibility and code reuse.

Recently, developers have liked many PHP frameworks, such as Symfony, CodeIgniter, Laravel, Laminas, Cake PHP, or Yii. Although they have many advantages, choosing the best PHP framework remains a challenge, and developers need to know how to choose the most appropriate one with the best support features to use in their project [12]. The authors [14] compared and analyzed the effect of the data abstraction layer (ORM) on the performance of CakePHP and CodeIgniter. The result of the analysis was the fact that Cake PHP is suitable for enterprise and large applications, while CodeIgniter was suitable for small and medium-sized applications. François Zaninotto and Fabien Potencier [15] emphasized that the tools provided by Symfony are for a professional environment. The authors [16] state that the Laravel framework is more suitable for large web projects that require faster delivery with fewer resources. Several sources indicate that the most popular frameworks are Laravel, Symfony and CodeIgniter.

From the comparison of the three frames, it can be concluded that all three compared frames are at a similar level. Only the details decide the best of them, and Laravel paid the most attention to them. Symfony is rated slightly worse, but it is clearly competitive and popular with developers. However, the author [12] mentions that CodeIgniter no longer meets the requirements of the current web, because there are not enough developers who develop this framework, and therefore its use in the web community is declining. When we focus on the performance of individual frames, in terms of requests per second, memory usage, response time or file size, Laravel became the clear winner. The code in this framework is optimized and fast. The author [12] concludes that Laravel is suitable for anyone who builds highly scalable applications or has little experience with PHP frameworks, because it is easier to understand, such as Symfony. For advanced developers, there is Symfony, which provides them with comprehensive options, but at the cost of a worse learning curve for this framework. CodeIgniter is a lightweight framework that leaves the developer a lot of freedom. It is ideal for short and simple projects.

V. WEB-BASED SOLUTION: DESIGN AND IMPLEMENTATION

The purpose of this assignment was to create two web applications. The first concerns the internal telephone directory of employees of the Technical University in Košice (UVP TECHNICOM). It combines all contacts in one place, in a clear form and is easy to manage contacts.

Today, the development of a dynamic web is a matter of several languages and technologies, which we will gradually introduce. However, it is good to look at the website, which we are going to develop in terms of parts or modules (see Fig. 1).

Like most websites today, our solutions will consist of a client and a server part. In the part for the client, we will display the data, or from users, in this case it will be only administrators, collect and then send to the server part. Intuitive, modern design and also the speed of data display will be an important element in this part. The server part will take care of manipulating the data that will be stored by the database. In the case of classic users, it extracts the data from the database, processes it into the required form and sends it to the client part, which will take care

of its display. However, it will also take care of inserting new entries, either in the phone book or in the list of courses. Security, data validation and also speed will be important elements.

Based on the proposal, two solutions have been developed. Namely:

- Telephone directory of the University Science Park TECHNICOM.
- List of courses provided by the Technical University in Košice to Erasmus students.

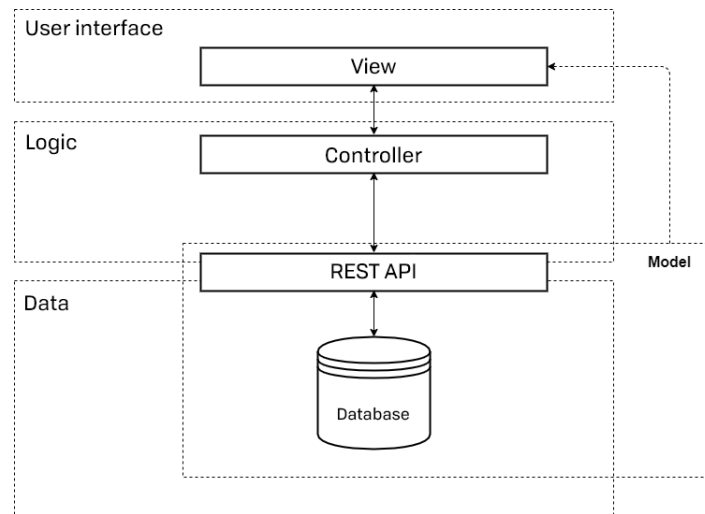


Fig. 1: Web application diagram

These solutions, or their server part, will be built on the same basis. This will be the Laravel PHP framework implementing the MVC model. In cooperation with MySQL, they will form the server part of the website. In the client part we will use a template system from Laravel - Blade, most of which consists of HTML markup language. The design will be in charge of the CSS language, which we will write in a CSS preprocessor called SASS. JavaScript in combination with the JQuery library, which contains useful features to make the programmer's life easier, will handle or send additional information while the page is running.

A. Preparation of data

Since the nature of the solution lies mostly in processing and displaying records, it is a logical step to collect all the available data that can be obtained. By analyzing the current solution and also communicating with a competent person, we found that the data is statically and hard-coded in HTML code, whether in the case of a telephone directory or a list of university courses. However, with today's browsers, it is possible to view the HTML structure in detail. In addition, we can run our own script written in JavaScript over the object model of the document (DOM), which, for example, we can obtain the available data, which is available in HTML code.

In both existing solutions, the data is structured regularly, so the way we can extract data from the page outside of manual rewriting is to write an algorithm that exports this data from the HTML code to JSON. We wrote such a script for both solutions, where the result was data in the form of JSON strings, which we then inserted into the tool for generating SQL INSERT from JSON strings.

Then all the necessary tables were created in the MySQL database, which we will use in the solutions in the solutions. We took SQL INSERTS generated from JSON objects and filled the tables with data.

B. Routing

Since the web application we create is divided into several parts and contains a system for managing the displayed data, we used the routing option, which is part of the Laravel framework. These paths are defined in a separate file by the methods that are designed to do so. We used POST and GET road types. GET paths are used for the classic opening of subpages in the browser. In contrast, POST paths are used to communicate between the client part and the server part during the run of the application, for example when searching or managing records. There is currently a combined administration interface for both solutions. For these paths, we used the prefix admin, under which all paths of the administration interface are defined and uses the authentication module (middleware). The path that enters the admin interface is defined as / admin. These defined paths then refer to the method defined in the specified driver, which processes the request and sends the response back to the client part.

C. Authentication

To ensure a quality solution to this problem, we used their own additional package based on Laravel's documentation - Laravel Fortify. From the point of view of the server part, it contains everything we need to run the authentication system.

The database takes care of storing individual users and other related data. It contains the user identifier, name, e-mail, password imprint (using the bcrypt algorithm) and a storage token. The registration system is implemented and available on the way / registration. In order for a registration not to be publicly available, it is conditional on the site operator allowing it in an environment variable named *AUTH_ENABLE_REGISTRATION*. With the login details, we can then log in on the path / admin or / login. If we click the Remember me option, an authorization token will be stored in the user's cookies, which will facilitate the login process.

D. Client part structure and design

In the client part we talk about the structure of HTML files, JavaScript code and styling using SASS and Bootstrap framework. HTML files are generated by the server side. Because the web interface is built on the principle of a one-page application, everything that renders and manipulates the data is taken care of by JavaScript, which is divided into several files.

Blade files contain all the content that is displayed in our solutions, whether it is a list of UVP employees or a list of TUKE items. One of the many functionalities of these files is that we can link them in different ways. Thus, we produce well-structured code, logically divided into folders and files. In this direction, the structuring style is used, where we have one main file (layout) and add content from other files to it.

When creating the website design, we tried to use the same colors and structure of the original solutions so that the new content fit into the original website. As for public websites, much of it was styled using SASS and seconded by Bootstrap. Otherwise, the goal for the administrator interface was to create a simple interface, where the primary styling tool was Bootstrap, with a small part customized using SASS and CSS, respectively. Responsiveness has been considered from the very beginning of development, so emphasis has been placed on adapting new web solutions to users on different devices.

E. Interface for administrators

The interface for administrators in the context of this solution is built on the same basis. Courses are divided into individual faculties displayed in a simple and intuitive interface (Fig. 2). After selecting the faculty, we can see the following part of the website. At the top we see the option to add or search for an item. In the lower part, you will find the option of switching the page on which we are currently located. We don't have to scroll with the mouse indefinitely to get to a certain object.

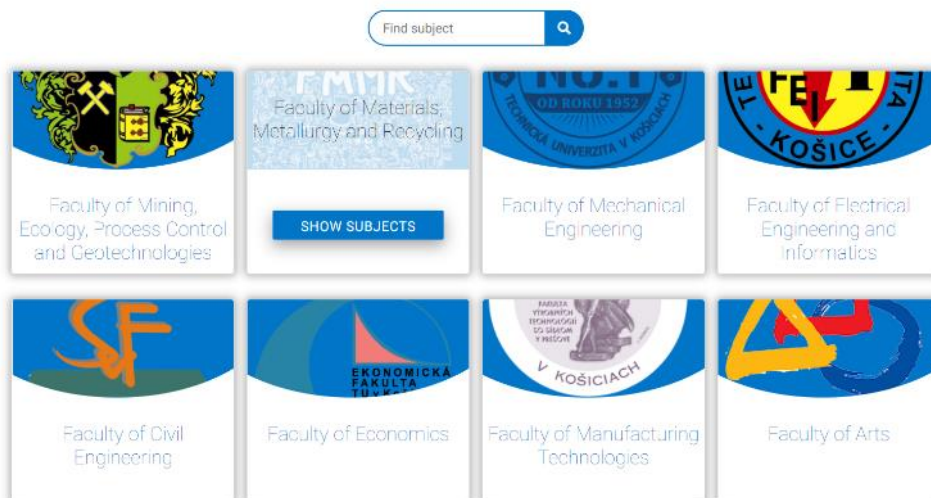


Fig. 2: Interface for the list of courses - faculties.

F. Use of web services

Much of the processing on the client part took the development of JavaScript (JS) and its interaction with either the web document or the server part. Within the public parts of our solutions, this scripting language was not used to the same extent as in the administration interface. One of the reasons is that we generate the content itself compared to the admin interface using Laravel Blades and PHP. Therefore, with the help of JS, we do not have to download information from the server part and process it.

The potential of web services is fully utilized in the administrator interface. When we open the admin interface for a particular faculty or contacts, the JS in the background loads the amount of data that is available. Based on this, it generates paging. It then downloads all relevant data in limited numbers. Subsequently, the JS instructs the function to be performed to all generated buttons within the contacts. This process is repeated each time you change the page number we are on. Searching for an item sends a request to the web service along with the search string, which asks the database itself to search. Subsequently, the web service returns a field of relevant data along with their number, so that we can create a paging page for many matches. Adding a contact or course works relatively easily. After clicking on the button, JS will first display a form where we enter the necessary information. It then sends them to a web service that takes care of adding the data to the database. At first glance, like adding a new contact, but still a little more complicated. If we want to edit a record, by clicking on the specified button JS will send a request to a web service that provides all the details of the record. If the web service returns JavaScript data, the web browser displays a form with pre-filled fields. If we subsequently edit the fields and click on "save", JS will collect the information from the displayed form and send it to the web service, which takes care of updating the records.

IV. RESULTS

The assignment was the development of a web solution for the management of records of the list of UVP TECHNICOM employees and the list of TUKE courses. In addition to a web-based interface for managing these records, another goal was to create a more intuitive interface for regular users.

Like every software solution, it was necessary to test and detect possible errors in functionality, security. Nowadays, web solutions need to work equally well on all types of devices, so the emphasis was on the responsiveness of the solution.

Based on testing, we can conclude that the application is ready for deployment and only unpredictable errors could occur. The tests were divided into user testing and, in addition, other tests, which focused on testing a specific part of the solution. We have been working with backed up production data since the beginning of development.

A. User testing

As part of testing the solution, the URL was sent to ten users. When choosing, we made sure that the properties of the users were evenly distributed. Selected users were from seventeen to forty-nine years old. Users who have no previous experience with similar systems, but also more advanced users, were selected.

One of the requirements was to test both public interfaces in the role of an ordinary visitor, where the testers had to test the intuitiveness, speed, responsiveness and functionality of the solution. The testers were not given scenarios but were given free hands and time to get into unexpected situations.

The second requirement was to test the interface for administrators. In this test, users used the solution as an administrator. So, they could manipulate the records and try out the complete functionality. The task was to test intuitiveness, functionality and responsiveness again.

After testing the solution, users had a questionnaire at their disposal, where they could evaluate the appropriate solution or report errors. The evaluation worked on the basis of a standardized System Usability Score. In practice, this means 1/5 (worst) to 5/5 (best). The questionnaire was required to be completed by all users who tested the solution.

B. Testing results

The test results show that the solution is in good condition. Regarding the interface for regular visitors, we conclude that three out of ten users rated the intuitiveness of the public website for the list of courses with four points out of five possible. Apart from this variation, users with a regular user interface had no problem, whether on large screens or smaller devices. In the admin interface, the user was missing a selection of several records at once and also one of the testers

reported that after clicking on the word "Admin panel" it would link to a page that was not expected.

In general, user testing can be considered successful. No functionality errors were reported, only details, or functionality that can be added to the solution.

V. CONCLUSION

In the applied research, the two web solutions have been developed and tested, both, consisting of an administrator interface and a public section for users. Solution may be deployed on university domains, thus allowing students or visitor to use them if necessary. When developing the solution, we first analyzed the web services that were used in the final solution. We also performed an analysis of CMS systems, their principles and architecture. Subsequently, we analyzed the PHP framework, based on which we chose the most suitable framework for our web application. Based on the mentioned analyzes, we gradually developed a system based on PHP under Laravel, where records management using the administrator interface works on the basis of web services.

One of the benefits is storage of records in a database. In the original solution, the data was stored statically in HTML code, where it was very difficult to manipulate, modify or delete it. The data is stored in a database with which the server part of the solution communicates and sends information from it to the client part. Another addition is the administration interface for managing records of both solutions. Thanks to the storage of data in the database, it was possible to create an administrator panel, where we can manage all existing records. Records can be added, deleted and edited. The new user interface for the TUKE course list is definitely another addition. A change in this solution was necessary due to poor usability of the original elements.

VI. ACKNOWLEDGMENT

This publication is the result of the Project implementation: University Science Park TECHNICOM for Innovation Applications Supported by Knowledge Technology, Phase II., ITMS: 313011D232, supported by the Research & Innovation Operational Programme funded by the ERDF. We support research activities in Slovakia/This project is being co-financed by the European Union.

VII. REFERENCES

- [1] NAKWASKI, M.; ZABIEROWSKI, W. Content management system for web portal. In: *2010 International Conference on Modern Problems of Radio Engineering, Telecommunications and Computer Science (TCSET)*. 2010, p. 233–23.
- [2] AUSTIN, Daniel; FERRIS, Christopher; BARBIR, Abbie; GARG, Sharad. *Web Services Architecture Requirements*. 2004-02. W3C Note. W3C. <https://www.w3.org/TR/2004/NOTE-wsa-reqs-20040211/>.
- [3] GOTTSCHALK, K.; GRAHAM, S.; KREGER, Heather; SNELL, J. Introduction to Web services architecture. *IBM Systems Journal*. 2002, p. 170–177. DOI: 10.1147/sj.412.0170.
- [4] PETERSEN, Sobah. *Web Services: Architectures and Standards*. 2020.
- [5] SUNYAEV, Ali. *Web Services*. In: 2020, s. 155–194. ISBN 978-3-030-34956-1. DOI: 10.1007/978-3-030-34957-8_6.
- [6] GRAHAM, Steve. *Building web services with Java : making sense of XML, SOAP, WSDL, and UDDI*. Indianapolis, Ind: Sams, 2004. ISBN 9780672326417.
- [7] CARDOSO, Jorge. *Semantic Web services : theory, tools, and applications*. Hershey, PA: Information Science Reference, 2007. ISBN 9781599040455.
- [8] CABRERA Kurt, Box. *Web services architecture and its specifications : essentials for understanding WS*. Redmond, Wash: Microsoft Press, 2005. ISBN 0735621624.
- [9] PAPAOGLOU, M. *Web services : principles and technology*. Harlow, England New York: Pearson/Prentice Hall, 2008. isbn 978-0321155559.
- [10] FRANKOWSKI, Paweł. *CMS : Jak szybko i łatwo stworzyć stronę WWW i zarządzać nią*. Gliwice: Wydawnictwo Helion, 2007. ISBN 83-246-0809-5.
- [11] MANOJ K. SRIVASTAV, A. Nath. *Web content management system*. 2016.
- [12] LAAZIRI, Majida; BENMOUSSA, Khaoula; KHOULJI, Samira; KERKEB, Mohamed. *A Comparative study of PHP frameworks performance*. *Procedia Manufacturing*. 2019, p. 864–871. DOI: 10.1016/j.promfg.2019.02.295.
- [13] OLANREWAJU, Rashidah; ISLAM, Thouhedul; ALI, Norashikin. *An Empirical Study of the Evolution of PHP MVC Framework*. In: 2015, p. 399–410. ISBN 978-3-319-07673-7. DOI: 10.1007/978-3-319-07674-4_40.
- [14] FAYYAZ, Ali; MUNIR, M. Performance Evaluation of PHP Frameworks (CakePHP and CodeIgniter) in relation to the Object-Relational Mapping, with respect to Load Testing. In: 2014.
- [15] PROKOFYEVA, Natalya; BOLTUNOVA, Victoria. Analysis and Practical Application of PHP Frameworks in Development of Web Information Systems. *Procedia Computer Science*. 2017, p. 51–56. DOI: 10.1016/j.procs.2017.01.059.
- [16] DAS, R.; SAIKIA, Dr. Lakshmi Prasad. Comparison of Procedural PHP with Codeigniter and Laravel Framework. In: 2016.

Neurónové siete typu RBF pri navigácii robota

Ján VAŠČÁK

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

jan.vascak@tuke.sk

Abstrakt – Tento článok sa zaoberá problematikou navigácie mobilného robota a simultánneho mapovania jeho okolia pomocou prístupu SLAM. Úlohou je zabezpečiť pohyb robota z ľubovoľného počiatočného bodu pomedzi rozmiestnené prekážky do stanoveného cieľa v predpísanom časovom limite, pričom samotná navigácia je založená na údajoch o vzdialenostiach k prekážkam a odhadoch prejdenej trajektórie na základe odometrie. V tejto práci bola využitá neurónová sieť typu RBF, ktorá je založená na princípoch hybridnej výpočtovej inteligencie. Pre overenie kvality navrhnutého riešenia boli vykonané simulačné experimenty v prostredí simulátora Webots.

Kľúčové slová – hybridná výpočtová inteligencia, mobilný robot, navigácia, RBF sieť

I. ÚVOD

Cieľom tohto článku je poukázať na možnosti využitia prostriedkov tzv. hybridnej výpočtovej inteligencie, ktorá je založená na paradigme spájania rôznych prvkov umelej inteligencie s cieľom vytvorenia prostriedku, ktorý by kompenzoval nedostatky jednotlivých prvkov a zároveň posilňoval ich silné stránky. V tomto prípade sme sa zamerali na špeciálny typ neurónových sietí, ktoré využívajú tzv. funkcie so stredovou (lúčovitou) súmernosťou, známe pod skratkou RBF (*Radial basis Function*).

Pre potreby otestovania navrhnutého prostriedku vytvoríme virtuálnu simuláciu, v ktorej bude mobilný robot navigovaný do stanoveného cieľa. Počas pohybu bude schopný sa autonómne vyhýbať okolitým prekážkam prostredníctvom riadenia založeného na RBF sieťach s využitím Gaussovej funkcie [1]. Táto sieť bude následne natrénovaná tak, aby bol robot schopný sa vyhnúť náhodným prekážkam v prostredí, ktoré bránia robotovi v jednoduchom a priamočiaram pohybe do cieľa. Riešenie by malo zahŕňať aj odometriu dvojkoľosového diferenciálneho podvozku robota a taktiež tzv. prístup SLAM, t. j. *simultánnu lokalizáciu a mapovanie* okolitého priestoru, čím vznikne mapa prostredia, v ktorom sa mobilný robot pohybuje [2].

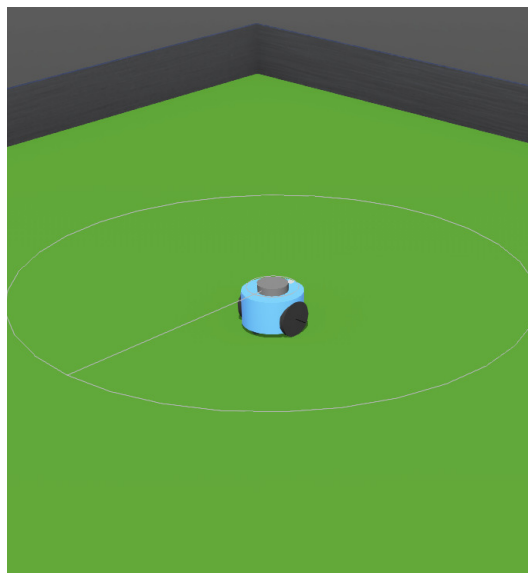
II. POPIS POUŽITÝCH PROSTRIEDKOV

Pre potreby experimentov bola zvolená síce jednoduchá konštrukcia mobilného robota, ale zabezpečujúca realistické vlastnosti podobné iným známym robotom valcovitého typu (kruhový pôdorys), ako sú napríklad roboty TurtleBot, Khepera alebo E-puck [3]. Tento robot bol kompletne vymodelovaný v simulátore Webots a disponuje týmito časťami, vid' Obr. 1:

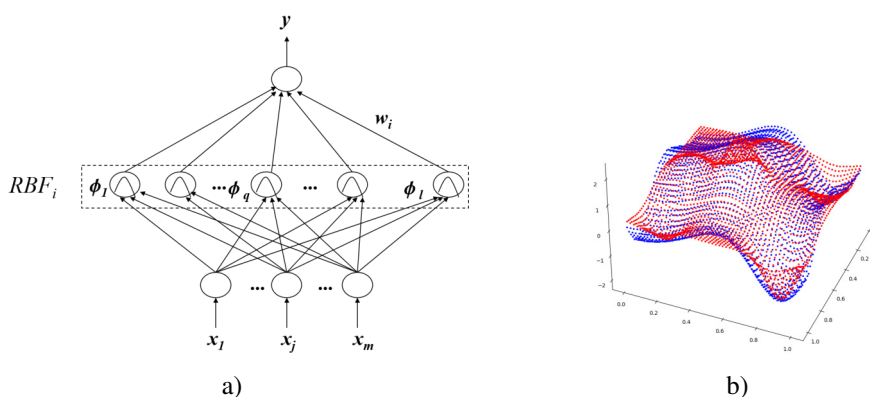
- dvojkolesovým diferenciálnym podvozkom s paralelným radením,
- dvomi motormi (rotačné kĺby),
- dvomi guľovými ložiskami (vpredu a vzadu),
- dvomi prevodníkmi (na oboch motoroch),
- jedným lidarom.

RBF sieť je matematický konštrukt hybridnej výpočtovej inteligencie, ktorý využíva ako aktívne funkcie v skrytej vrstve funkcie so stredovou súmernosťou, najčastejšie Gaussovského typu, vid' Obr. 2a. Funkciám v skrytej vrstve je potrebné pred začatím procesu učenia inicializovať hodnoty distribúcií a stredných hodnôt. Na to je možné použiť rôzne metódy strojového učenia založené na nekontrolovanom učení ako napríklad *K-Means*. Cieľom je nájsť zhľady v súbore tréningových dát, ktoré predstavujú vhodné stredné hodnoty a distribúcie ako parametre týchto funkcií.

Pri doprednom reťazení je na vstupnej vrstve vektor nameraných hodnôt predstavujúci vstup do skrytej vrstvy RB funkcií, kde sú vypočítané hodnoty aktivácie. Výstupný vektor z funkcií je lineárnou kombináciou ich funkčných hodnôt a váh smerujúcich k výstupnému neurónu, vid'



Obr. 1 Mobilný robot s lidarom v simulačnom prostredí Webots.



Obr. 2 Ukážka a) štruktúry a b) 3D vizualizácie výstupu RBF siete.

Obr. 2b. Váhy spojení smerujúce k výstupnému neurónu sú pred začatím procesu učenia inicializované na náhodné hodnoty. V procese učenia sa na výstupe po ukončení dopredného reťazenia určí chyba, ktorá je vstupom pre chybovú funkciu. Modifikovaným algoritmom spätného šírenia chyby sú následne upravované váhy medzi skrytou a výstupnou vrstvou.

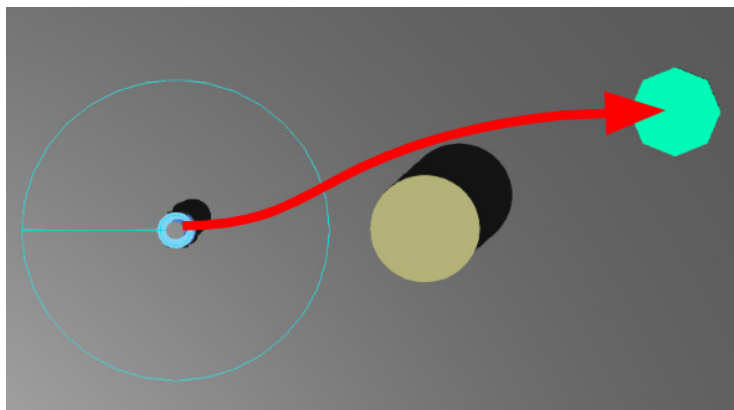
Použitie RBF sietí je rovnaké ako u klasických neurónových sietí, t. j. aproximácia ľubovoľnej po častiach diferencovateľnej funkcie, predikcia časových radov alebo klasifikácia vstupov do tried. Podobne je to aj s tréňovaním, keď je potrebné mať už existujúce vstupy a k nim prislúchajúce výstupy, ktoré budú slúžiť ako tréňovacia množina [4]. V našom prípade bude neurónová sieť vykonávať regresnú úlohu, ktorej vstupom budú dáta z lidar a výstupom bude pomer uhlových rýchlostí motorov.

III. NÁVRH RIEŠENIA

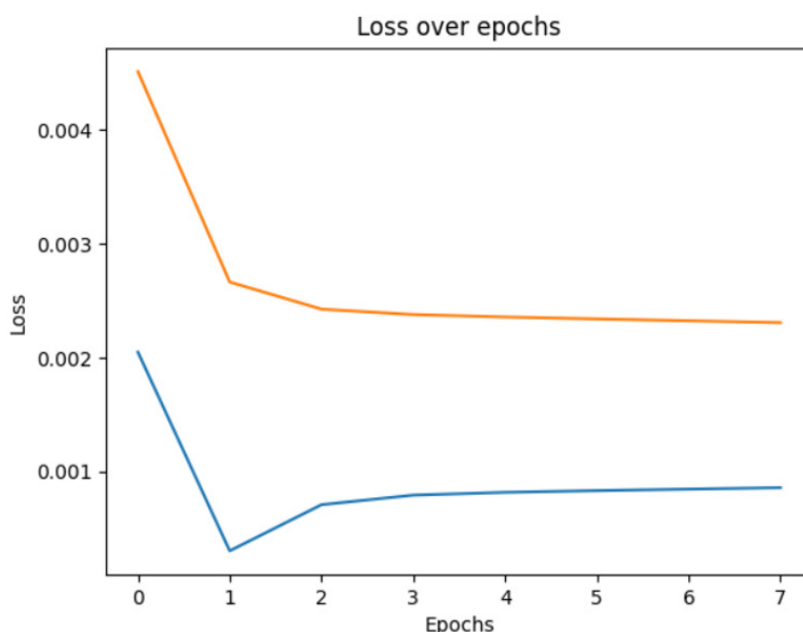
V tejto kapitole si bližšie preberieme a popíšeme jednotlivé časti nášho riešenia navigácie mobilného robota. Taktiež sa tu nachádza popis algoritmu navigácie robota riadeného pomocou RBF siete.

Prvým krokom je získanie súboru vhodných dát, ktoré by podľa možností dostatočne reprezentatívne pokrývali čo najväčšiu časť stavového priestoru. Takýto dátový súbor obsahuje údaje z lidar o nameranej vzdialenosti, konkrétne z deviatich jeho lúčov pokrývajúcich interval od -90° po $+90^\circ$.

Dátový súbor je štruktúrovaný do formátu .csv a obsahuje približne 1000 záznamov z manuálneho riadenia robota v simulačnom prostredí pri rôznych pohybových úlohách, vrátane vyhýbania sa prekážkam (Obr. 3), ktorý bol následne použitý pri tréňovaní siete. Tieto dáta popisujú snímanie lidar počas manuálneho ovládania robota, kde posledný stĺpec popisuje uhol otočenia robota a môže byť z rozsahu od 0° po 360° .



Obr. 3 Ukážka vyhýbania sa robota prekážkam v prostredí Webots.



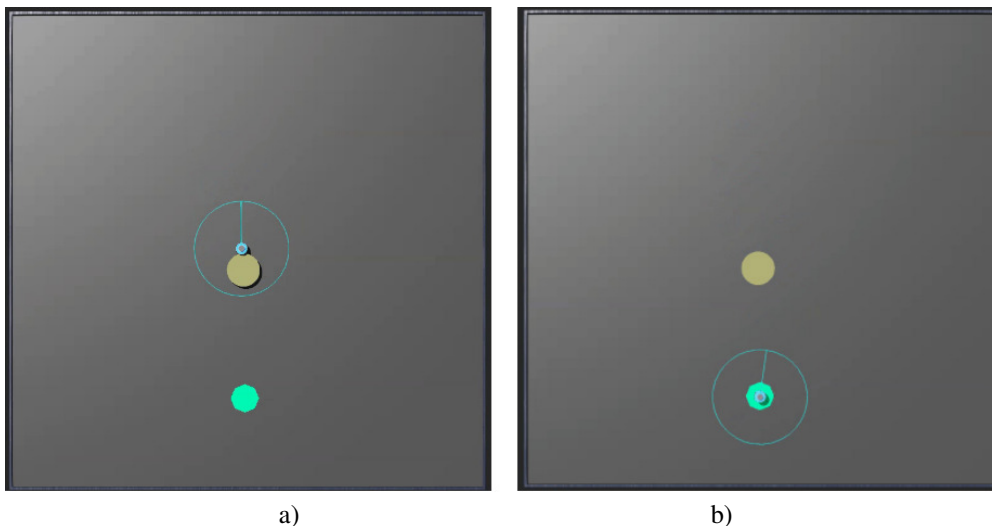
Obr. 4 Porovnávací graf tréningu *preučeného* (modrý priebeh) a dobre naučeného (žltý priebeh) modelu.

Proces samotného tréningu siete sa skladá z dvoch častí. Prvá časť zahŕňa tréning ľavo-točivého modelu a druhá časť tréning pravotočivého modelu. Dôvodom pre takýto prístup je hlavne zistenie, že jeden model sa vie natréňovať len na jeden typ vyhýbania sa. Teda buď sa RBF sieť natrénuje na vyhýbanie sa doľava alebo vyhýbanie sa doprava. Touto metódou je eliminované vyhýbanie sa prekážkam do dvoch separátnych modelov, čím je zabezpečená všeobecne vyššia spoľahlivosť navigácie robota.

V procese učenia bola zaznamenaná hodnota chybovej funkcie modelu v jednotlivých epochách do grafu, kde hľadáme minimálnu tzv. stratovú hodnotu. Na Obr. 4 je ukázané tréning na ľavotočivom modeli (tréning na pravotočivom modeli prebieha obdobným spôsobom). Modrý graf popisuje tzv. *preučený* model s malou chybou na tréningových dátach po ôsmich epochách, čo však na ostatných dátach zapríčinilo nesprávne natréningovanie a nefunkčnosť tohto modelu. Žltý graf popisuje dobre naučený model, so síce vyššou chybou na tréningových dátach po ôsmich epochách, avšak pre ostatné dáta bol tento model použiteľný na rozdiel od toho *preučeného*.

Oba spomenuté modely boli otestované po procese tréningu na rovnakej mape simulačného prostredia, aby sa zistilo správanie modelov, vid' Obr. 5. Pri tzv. *preučenom* modeli (Obr. 5a) sa robot správal tak, že prekážku odignoroval a do nej narazil. Pri dobre naučenom modeli (Obr. 5b) však robot prekážku obišiel z jednej strany (v závislosti od toho, do akej strany bol model natréňovaný) a po jej obídení pokračoval vo svojej ceste za cieľom.

V procese testovania natréňovaných modelov sa vyhodnocovala spoľahlivosť navigácie robota pri obchádzaní prekážok do cieľa. Natréňovali a nasadili sa obidva (ľavotočivý a pravotočivý)



Obr. 5 Porovnanie správania sa a) *preučeného* a b) dobre naučeného modelu.



Obr. 6 Testovanie na jednoduchých mapách.

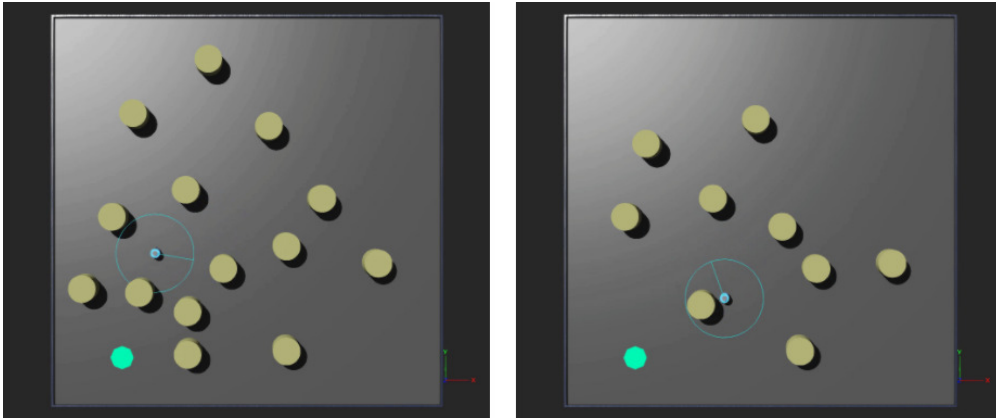
modely, čím sa zabezpečila univerzálnosť vyhýbania. Mobilný robot prepínal medzi týmito modelmi na základe senzorickeho vstupu z lidar, pričom päť lúčov sa nastavilo na jeden typ modelu a zvyšné štyri na druhý typ modelu. Najprv sa riešenie otestovalo na jednoduchých mapách s jedinou prekážkou medzi počítačovou pozíciou robota a cieľom, vid' Obr. 6.

Po zapnutí simulácie je algoritmus robota naprogramovaný tak, aby sa priamočiaro dostal robot do cieľa. Pri detekcii okolitej prekážky robot preruší svoj priamočiary pohyb a začne predikovať číselnú hodnotu pomeru rýchlostí kolies. Po vyhnutí sa prekážke je predikcia tejto hodnoty prerušená a robot opäť vykonáva priamočiary pohyb do cieľa. Následne boli použité mapy so zložitejším rozmiestnením prekážok a testovali sa obidva modely z hľadiska miery ich spoľahlivosti a univerzálnosti, vid' Obr. 7.

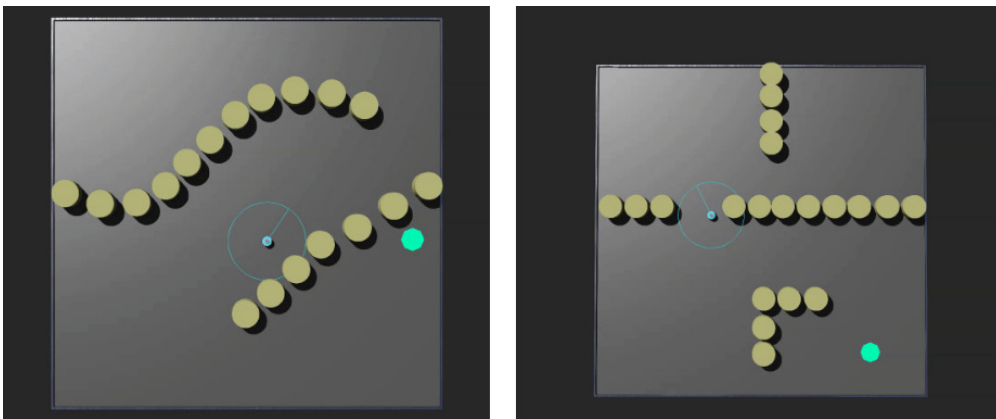
Pokusy ukázali, že mobilný robot bol skutočne schopný sa presunúť zo svojho počítačového stanoviska do konečného cieľa za pomoci RBF navigácie, ak lidar zaznamenal nejakú prekážku v aspoň jednom zo svojich lúčov. Na ilustráciu univerzálnosti navrhnutého riešenia boli vytvorené aj špeciálne mapy v podobe dvoch bludísk, kde robot v jednom kvadrante sa musí dostať do cieľa umiestneného v inom kvadrante, Obr. 8.

IV. EXPERIMENTY A ICH VYHODNOTENIE

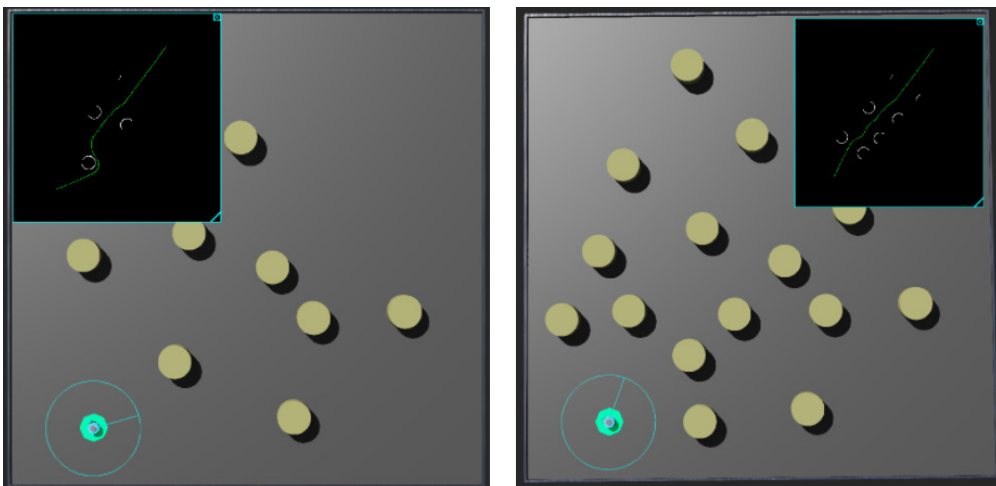
Experimenty boli vykonané na rôznych mapách s rôznou obtiažnosťou, pričom sa líšili rozlohou, počtom prekážok a ich usporiadaním, počítačovou pozíciou a natočením robota, či rôznou polohou cieľa, pričom bola sledovaná schopnosť robota sa dostať do požadovaného cieľa pomocou natrénovaného modelu. Zároveň popri navigácii robota bolo vykresľované aj okolité prostredie do diskretnej obrazovej roviny, kde sa okrem trajektórie robota (zelená farba) vykresľovali aj okolité prekážky (biela farba), Obr. 9.



Obr. 7 Testovanie na zložitejších mapách.



Obr. 8 Testovanie na bludiskách.



Obr. 9 Vykresľovanie okolitého prostredia a trajektórie.

V. ZÁVER

Využitím navigačnej RBF siete sa podarilo úspešne nasimulovať mobilného robota, ktorý je schopný sa pohybovať v priestore virtuálneho sveta v simulačnom prostredí Webots. Na základe odometrie bola určovaná poloha mobilného robota a na mapovanie okolia robota bol použitý lidar. Navrhnuté riešenie je použiteľné v širokom spektre rôznych druhov aplikácií vzhľadom na univerzálnosť vytvoreného algoritmu. Takýto aparát je zároveň aplikovateľný aj na reálneho robota.

Toto riešenie je možné rôzne rozšíriť a vylepšovať ako napríklad implementáciou navigácie

cez viacero bodov. To by znamenalo navigáciu robota v priestore pomocou tzv. smerovníkov (waypoints), a tým sa dostať do cieľa cez konkrétnu trajektóriu. Ďalej by bolo možné upraviť modely RBF siete napr. opätovným natréňovaním modelov, čím by sa zvýšila presnosť navigácie.

POĎAKOVANIE

Táto publikácia je výstupom projektu ENISaC – Edge-enabled inteligentné snímanie a výpočty (APVV-20-0247), podporovaného Agentúrou na podporu vedy a výskumu (APVV).

LITERATÚRA

- [1] S. Ishikawa, "A method of indoor mobile robot navigation by using fuzzy control," in *Proceedings IROS '91:IEEE/RSJ International Workshop on Intelligent Robots and Systems '91*, 1991, pp. 1013–1018 vol.2.
- [2] P. Gyawali and P. K. Agarwal, "Fuzzy behaviour based mobile robot navigation in static environment," in *2018 IEEE Recent Advances in Intelligent Computational Systems (RAICS)*, 2018, pp. 190–194.
- [3] O. Saha, P. Dasgupta, and B. Woosley, "Real-time robot path planning from simple to complex obstacle patterns via transfer learning of options," *Autonomous Robots*, vol. 43, no. 8, pp. 2071–2093, 2019.
- [4] P. K. Panigrahi, S. Ghosh, and D. R. Parhi, "Intelligent leaning and control of autonomous mobile robot using mlp and rbf based neural network in clustered environment," *International Journal of Scientific & Engineering Research*, vol. 5, no. 6, 2014.

Návrh 0,32 Tb/s Ultra-DWDM systém s optickým zosilňovačom ROA

¹ Norbert ZDRAVECKÝ, ² Luboš OVSEŇÍK, ³ Samuel ANDREJČÍK, ⁴ Jakub ORAVEC, ⁵ Maroš LAPČÁK

^{1,2,3,4,5} Katedra elektroniky a multimediálnych telekomunikácií, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

¹norb.zdravecky@tuke.sk, ²lubos.ovsenik@tuke.sk, ³samuel.andrejcik@tuke.sk, ⁴jakub.oravec@tuke.sk, ⁵maros.lapcak@tuke.sk

Abstrakt — Príspevok sa zaoberá návrhom a analýzou 0,32 Tb/s Ultra-DWDM systému s použitím Ramanovho zosilňovača (ROA). Návrh skúma zosilnenie jednotlivých kanálov pri použití ROA v Ultra-DWDM. Pozoruje tiež nežiaduce vplyvy nelineárnych javov, ktoré sú hlavným faktorom ovplyvňujúcim kapacitu a prenosovú vzdialenosť optického komunikačného systému.

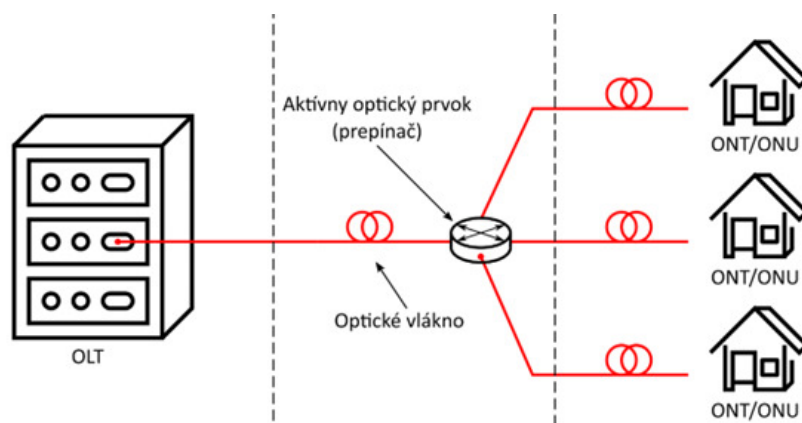
Kľúčové slová — BER, DWDM, nelineárne javy, ROA, ultra-DWDM

I. ÚVOD

Na využitie šírky pásma sú potrebné rôzne formy multiplexovania. Najbežnejším typom multiplexovania je multiplexovanie s časovým delením (TDM Time Division Multiplexing). Optické vlákno je najvhodnejšie médium na vysokorýchlostný prenos. Multiplexovanie s delením vlnovej dĺžky (WDM) má mnoho typov, napríklad CWDM, DWDM, Ultra -DWDM [1,4,7]. Optické zosilňovače sú kľúčovými zariadeniami v optických komunikačných systémoch. Ramanove zosilňovače boli skúmané ako alternatíva ku Erbium dopovaným zosilňovačom (EDFA) na realizáciu Ultra-DWDM systémov, pretože sú prirodzene širokopásmové a schopné poskytnúť zisk pri akejkoľvek vlnovej dĺžke [2,8]. Najdôležitejšie je, že sú flexibilné na tvarovanie profilu zisku správnym nastavením výkonu pumpy a vlnových dĺžok pri prevádzke v konfigurácii s viacerými pumpami. Toto je kritická vlastnosť pre budúce ultraširokopásmové systémy, pretože výkonové profily kanálov, ktoré maximalizujú dosiahnuteľnú informačnú rýchlosť, nemusia byť nevyhnutne ploché v dôsledku zvýšených úrovní Kerrovej nelinearity a stimulovaného Ramanovho rozptylu [3].

II. AKTÍVNE OPTICKÉ SIETE

Aktívne optické prístupové siete AON (Active Optical Network) sa odlišujú od pasívnych optických sietí v tom, že na prenos signálu využívajú okrem vysielača a prijímača aj ďalší aktívny optický prvok. Medzi aktívne prvky zaradujeme sieťové komponenty, ktoré potrebujú pre fungovanie elektrické napájanie. Najpoužívanějšími aktívnymi prvkami v AON sieťach sú zosilňovače optického signálu [5].



Obr. 1 Aktívna optická sieť

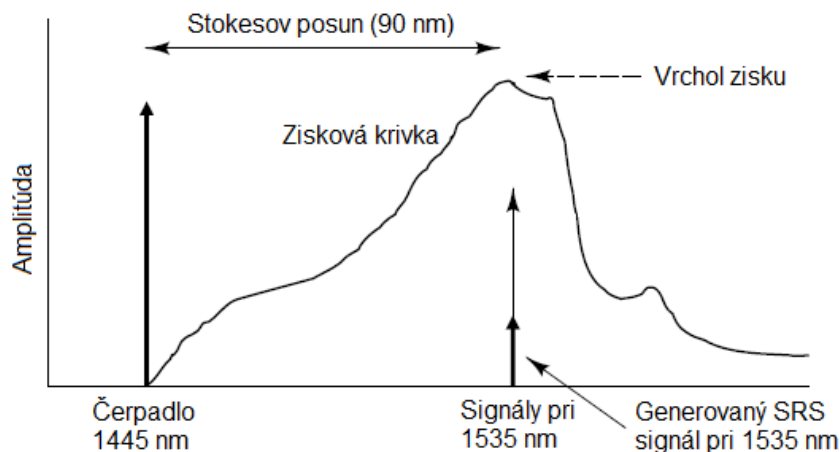
Hlavnou výhodou AON sietí je prenos signálu na väčšie vzdialenosti vďaka zosilňovačom a vysokej flexibilnosti aktívnych komponentov. Z dlhodobého hľadiska však nie je napájanie aktívnych optických komponentov finančne výhodné [6]. Na Obr. 1 môžeme vidieť optické linkové zakončenie OLT (Optical Line Termination), aktívny prvok, ktorý je v tomto prípade prepínač a optickú sieťovú jednotku ONU (Optical Network Unit).

III. RAMANOV ZOSILŇOVAČ

Ramanove zosilňovače ROA (Raman Optical Amplifiers) umožnili výrazné zvýšenie dosahu a kapacity. Optické zosilnenie je prezentované z nelineárných vplyvov vo vnútri vlákna. Prevádzkou stimulovaného rozptylu SRS (Stimulated Raman Scattering) možno tieto typy zosilnenia dosiahnuť zmenou výkonu optického signálu. SRS je spôsobený interakciou medzi optickým energetickým poľom a vibračnými režimami mriežkovej štruktúry v materiáli [2,8].

Zosilňovacia schopnosť ROA pokrýva veľkú šírku pásma, siaha medzi 1280 až 1650 nm, Ramanove zosilnenie signálov je teda možné vykonať v pásme C a L, kde sú potrebné čerpacie lasery s vysokým výstupným výkonom [8].

Stimulovaný Ramanov rozptyl vzniká ak atóm najskôr absorbuje fotón s určitou energiou a potom uvoľní ďalší fotón s nižšou energiou, čo znamená, že má dlhšiu vlnovú dĺžku. Energetický rozdiel medzi absorbovanými a uvoľnenými fotónmi sa transformuje na fonón. Signál generovaný SRS môže pôsobiť ako zámerné zosilnenie určitej vlnovej dĺžky alebo môže ísť o nežiaduci interferenčný signál [9].

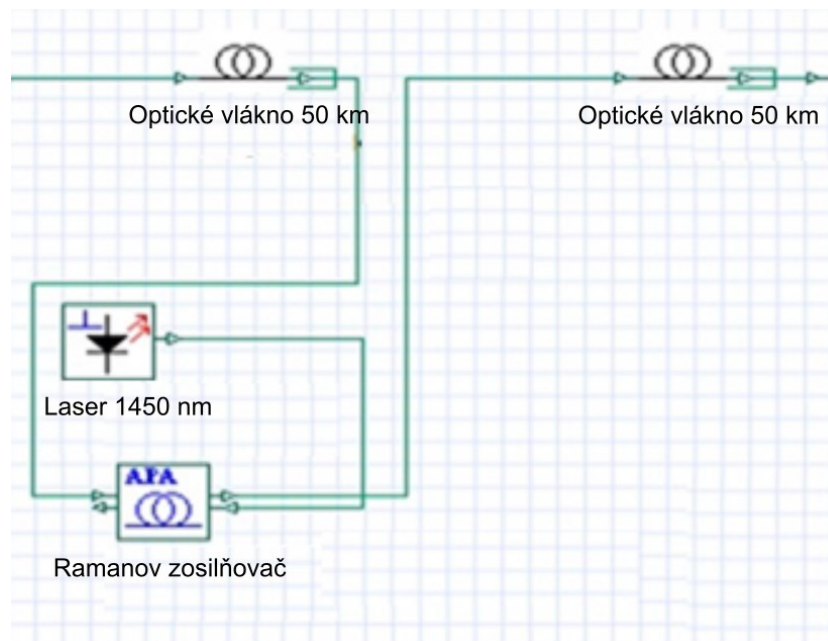


Obr. 2 Stokesov posun a výsledný Ramanov zisk z čerpadla pri 1445 nm

ROA využíva ako zosilňovacie médium štandardné optické prenosové vlákno. Poznáme distribuovaný ROA a sústredný ROA, ktorý je zložený z 80 m cievky optického vlákna a čerpacími lasermi vloženými do prenosovej cesty. Distribuovaný ROA nazývame vtedy ak optická energia z laserov Ramanovho čerpadla je vložená do prijímacieho konca prenosového vlákna. Tento proces premení posledných 20 až 40 km vlákna na predzosilňovač [8,9]. Jav SRS postupne prenáša energiu z kratších vlnových dĺžok čerpadla na dlhšie vlnové dĺžky signálu.

IV. NÁVRH -DWDM SYSTÉMU S OPTICKÝM ZOSILŇOVAČOM ROA

Návrh nášho 0,32 Tb/s DWDM systému vychádza z dlhotrvajúceho výskumu a množstva simulácií. Prenosová rýchlosť je 10 Gb/s na 1 kanál. Dĺžka prenosovej trasy tvorenej jednovidovým optickým vláknom je 50 km a dĺžka vlákna kompenzujúce disperziu je 10 km. Výkon lasera sa menil od -3 dBm po 3 dBm. Frekvencia dopovacieho lasera je 1450 nm. Prenosová časť navrhnutého systému je na Obr. 3. Ďalšie parametre navrhnutého modelu sú uvedené v Tab. 1. Analyzoval sa výkon na výstupe pre tri hodnoty vstupného výkonu.



Obr. 3 Prenosová časť Ultra-DWDM systému.

Optické vlákno s dĺžkou 50 km má Q faktor najvyšší pre hodnotu výkonu lasera 3 dBm. Rozdiel hodnôt Q faktoru pri výkonoch 0 dBm a 3 dBm je malý. Najnižšia hodnota je pri výkone lasera -3 dBm. Výsledky budú zobrazené v tabuľkách, rozdelené podľa čísla kanála a výkonu lasera. Tabuľka 1 zobrazuje stabilne nastavené hodnoty pre celý navrhovaný systém.

Tabuľka 1 Parametre navrhnutého Ultra-DWDM systému

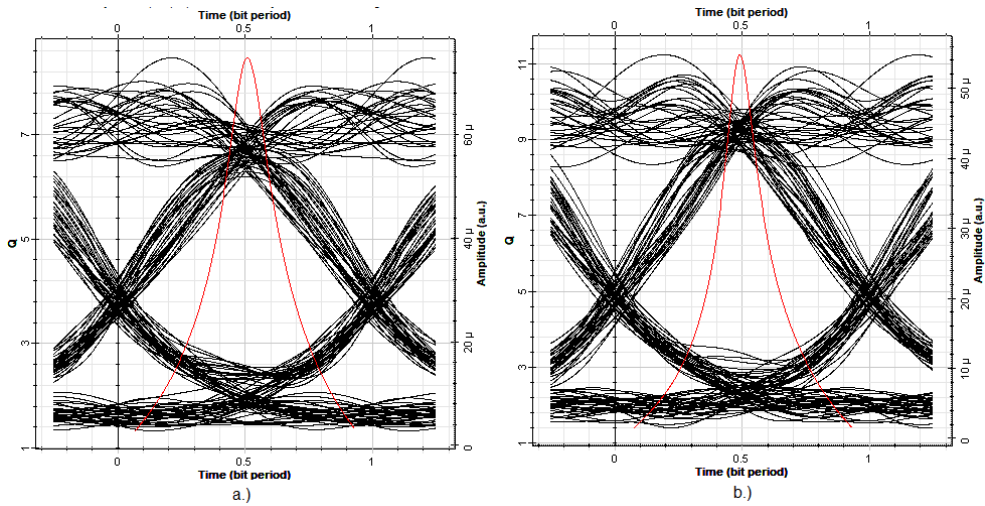
Vysielač – 32 kanálov	
Bitová rýchlosť	10 Gb/s
Počet vzoriek na 1 bit	64
Počiatočná frekvencia	193,1 THz
Medzikanálové rozostupy	100 GHz
	-3 dBm
Vyžarovací výkon	0 dBm
	3 dBm
Prenosová trasa	
Dĺžka SMF	50 km
Hodnota chromatickej disperzie SMF	17 ps/nm/km
Útlm	0,2 dB/km
Disperzia - β_2	-20 ps ² /km
Polarizačná disperzia – PMD	0,2 ps/km
Hodnota chromatickej disperzie DCF	-85 ps/nm/km
Dĺžka DCF	10 km

Limit BER pre DWDM systémy súčasnosti sa pohybuje okolo $1e-10$. Naše výsledky simulácii sa pohybujú na hodnotách $1e-17$, čo značne prevyšuje štandardy súčasných systémov.

Tabuľka 2 Výsledky Ultra-DWDM pre výkon lasera -3 dBm.

DWDM kanál	BER	Q faktor
CW výkon -3 dBm		
5	$1,25 \times 10^{-17}$	8,468
15	$1,02 \times 10^{-23}$	9,970
21	$2,58 \times 10^{-22}$	9,645
29	$1,29 \times 10^{-29}$	11,239

Výsledky Q faktora pri výkone lasera -3 sú zobrazené v Tabuľka 2 spolu s hodnotami BER pre jednotlivé kanály, najlepšia hodnota Q faktora je pri 29 kanály. Najnižšia hodnota pri nezmenených parametroch je na 5 kanály.

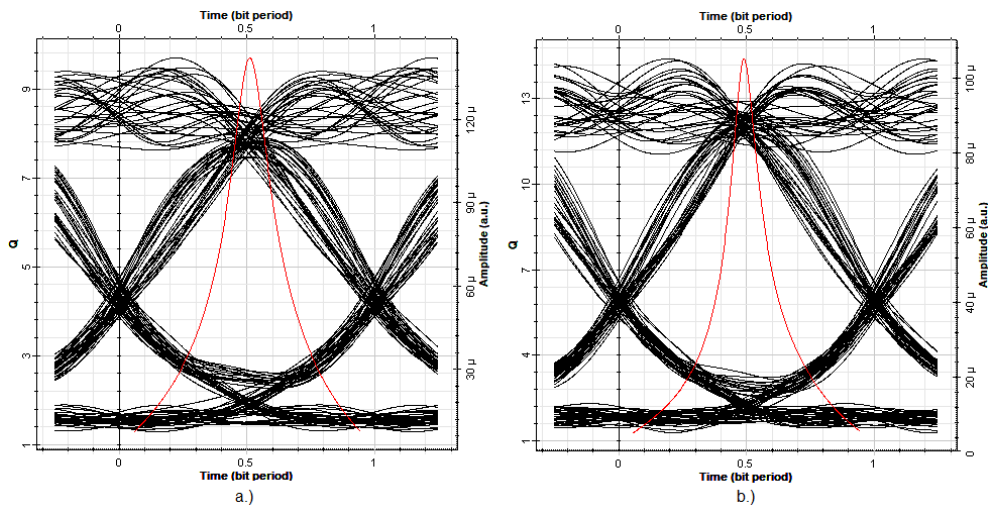


Obr. 4 Diagramy oka s využitím ROA zosilňovača pri výkone -3 dBm pre a.) 5. kanál, b.) 29. kanál

Tabuľka 3 Výsledky Ultra-DWDM pre výkon lasera 0 dBm

DWDM kanál	BER	Q faktor
CW výkon 0 dBm		
5	$1,58 \times 10^{-22}$	9,871
15	$3,96 \times 10^{-31}$	11,543
21	$5,28 \times 10^{-30}$	11,319
29	$2,64 \times 10^{-47}$	14,397

Ku každej simulácii sme vykreslili aj diagramy oka, ktoré nám pomáhajú s vizualizáciou výsledkov. Na diagrame oka môžeme sledovať viaceré parametre.

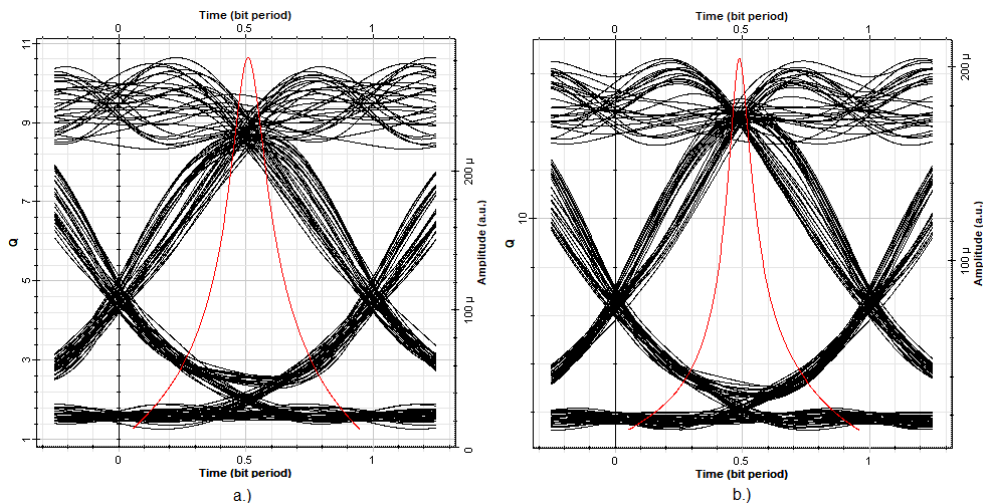


Obr. 5 Diagramy oka s využitím ROA zosilňovača pri výkone 0 dBm pre a.) 5. kanál, b.) 29. Kanál

Tabuľka 4 Výsledky Ultra-DWDM pre výkon lasera 3 dBm

DWDM kanál	BER	Q faktor
CW výkon 3 dBm		
5	$1,10 \times 10^{-26}$	10,633
15	$2,57 \times 10^{-40}$	13,239
21	$3,99 \times 10^{-37}$	12,677
29	$2,12 \times 10^{-62}$	16,639

Z Tabuľka 4 môžeme vyčítať, že najlepšia hodnota Q faktoru je pre kanál 29 a najnižšia pre kanál 5. Na Obr. 6 môžeme vidieť diagram oka pre kanál 5 a pre kanál 29. Červená čiara v diagrame zobrazuje BER.



Obr. 6 Diagramy oka s využitím ROA zosilňovača pri výkone 3 dBm pre a.) 5. kanál, b.) 29. Kanál

Ak by sme porovnávali parametre pri jednotlivých výkonoch lasera, tak pri -3 dBm hodnoty chybovosti BER boli na úrovni $1,25 \times 10^{-17}$ až $1,29 \times 10^{-29}$. Pri narastajúcom výkone lasera na 3 dBm sa hodnoty BER pohybovali $1,1 \times 10^{-26}$ až $2,12 \times 10^{-62}$. Pri analýze systému sme pozorovali ako sa zmení celkový výkon na výstupe signálu pre tri rôzne hodnoty výkonu lasera pre ROA zosilňovač.

V. ZÁVER

Bol navrhnutý Ultra-DWDM systém s využitím optického zosilňovača ROA v programe OptiSystem. Využilo sa jednovidové vlákno v dĺžke 50 km a vlákno potlačujúce disperziu v dĺžke 10 km. Hodnoty chybovosti BER a Q faktoru sú nad prahovými hodnotami. Výskum v oblasti zosilňovačov je dôležitý a umožňuje dosahovať stále vyššie prenosové vzdialenosti DWDM systémov. Práca v budúcnosti pomôže pri nasadzovaní jednotlivých typov zosilňovačov v Ultra-DWDM systémoch.

Schéma návrhu optickej siete so zapojením ROA zosilňovača sa vyznačuje širokým pásmom vlnových dĺžok oproti návrhu optickej siete so zapojením iného zosilňovača. Na základe dosiahnutých výsledkov možno konštatovať, že zosilňovače ROA je vhodné nasadzovať pri 10 Gb/s DWDM systémoch.

POĎAKOVANIE

Táto práca bola podporená výskumným projektom FEI-2022-84 "Data Processing Techniques in High-Speed Transmission Systems".

LITERATÚRA

- [1] Gerd Keiser: Fiber Optic Communication. USA: Newton Center, MA, 2021, ISBN: 978-981-33-4665-9, 639 pp.
- [2] Agrawal, Govind P.: Fiber-Optic Communication Systems (Fifth Edition). New York: The Institute of Optics University of Rochester, 2021, ISBN: 9781119737377, 544 pp
- [3] Gang-Ding, Peng: Handbook of Optical Fibers. 2019. University of New South Wales. ISBN: 978-981-10-7087-7, 2396 pp.
- [4] ITU-T G.694.1. Series G: Transmission Systems and Media, Digital Systems and Networks: Transmission media and optical systems characteristics – Characteristics of optical systems. Spectral grids for WDM applications: DWDM frequency grid. 2012.
- [5] Edward Harstead a Pieter H. Van Heyningen. Optical Access Networks. Optical Fiber Telecommunications IV-B. Elsevier, 2002, 2002, s. 438-513. ISBN 9780123951731. doi:10.1016/B978-012395173-1/50010-4
- [6] Elaine Wong, Maluge P. Imali Dias, Zhengxuan LI a Lilin YI. Broadband Passive Optical Access Networks. Encyclopaedia of Modern Optics. Elsevier, 2018, 2018, s. 73-86. ISBN 9780128149829. doi:10.1016/B978-0-12-803581-8.10107-9
- [7] A. Dochhan, at al., "Solutions for 80 km DWDM Systems", *Journal of Lightwave Technology*, 2015, pp. 491–499.
- [8] Vartan, V. Ter-Mikirtychev: Fundamentals of Fiber Lasers and Fiber Amplifiers (Second Edition). Florida Atlantic University. 2019. ISBN: 978-3-030-33890-9, 330 pp.
- [9] S. Kheraliya, Ch. Kumar: Comparative Study of Various Optical Amplifiers for 32-Channel WDM System. *Journal of Optical Communications*. 2018.

Model D2D Komunikácie pre 5G Sieť

¹Maros BAUMGARTNER, ²Jan PAPAJ, ³Erik SARSALA

^{1,2,3}Katedra elektroniky a multimediálnych telekomunikácií, Fakulta Elektrotechniky a Informatiky, Technická univerzita v Košiciach, Slovenská republika

¹maros.baumgartner@tuke.sk, ²jan.papaj@tuke.sk, ³erik.sarsala@student.tuke.sk

Abstrakt — Tento článok sa zameriava na použitie Device-to-Device komunikácie v mobilných bezdrôtových sieťach 5G a návrhom simulácie Device-to-Device komunikácie. Základom je stručný popis siete 5G a prípady použitia týchto sietí. Rozdelenie mobilných bezdrôtových sietí, kde spadá ďalej podrobne opísaná Device-to-Device komunikácia, jej rozdelenie podľa použitia spektra, možnosti využitia Device-to-Device komunikácie pre rôzne sféry, jej technické výzvy, ktoré je potrebné vyriešiť pred implementáciou. Opis technológie Blockchain a jeho využitie v informačných technológiách. Návrh simulácie Device-to-Device komunikácie v programovacom jazyku Python s implementáciou metód pre zvýšenie robustnosti komunikácie a vyhodnotenie výsledkov pomocou grafov a tabuliek.

Keľúčové slová — Blockchain, Device-to-Device komunikácia, 5G, bezpečnosť, dôveryhodnosť

I. ÚVOD

Neustálym rozvojom informačných technológií a nárastom zariadení, ktoré vyžadujú pripojenie do internetu sa sieť zahľucuje. Neustále zmenšovanie rádiových zdrojov pre jednotlivé zariadenia, znižujú kvalitu služieb. Príchodom novej generácie mobilných bezdrôtových sietí sa tieto problémy vyriešia a skvalitnia sa mobilné služby. Spolu s novou generáciou sietí, ktorá sprístupní aj doposiaľ nevyužívané spektrum nad 6 GHz, prinesie aj množstvo nových technológií, ktoré skvalitnia mobilné služby. Sieť piatej generácie, majú byť totižto siete všetkého.

Príchodom nových ale aj vylepšených súčasných technológií v sieťach piatej generácie je aj technológia Device-to-Device komunikácie. Technológia spočíva v priamej komunikácii medzi zariadeniami, napr. výmena dát medzi zariadeniami, bez toho aby dátová prevádzka smerovala cez základňovú stanicu. Takáto priama komunikácia zníži dátovú prevádzku na základňovej stanici, čím zlepši využitie jej rádiových zdrojov a týmto základňová stanica dokáže obslužiť väčší počet zariadení. Technológia Device-to-Device komunikácie nájde uplatnenie aj pri priamej komunikácii vozidiel, Vehicle-to-Vehicle, a aj pri priamej komunikácii strojov, Machine-to-Machine.

Stále viac zaujímavou novšou technológiou Blockchainu sa venuje vo výskume veľa pozornosti. Hlavnými výhodami tejto technológie je decentralizovanosť, takže nemá centrálny uzol, ktorý riadi celú technológiu, tým nie je technológia náchylná na útok. Blockchain funguje ako veľká neustále sa zväčšujúca databáza, ktorá je verejná a môže do nej nazrieť ktokoľvek.

V praktickej časti, bolo účelom navrhnuť simuláciu viacpreskokovej Device-to-Device komunikácie v programovacom jazyku Python. Implementácia metód na zvýšenie robustnosti viacpreskokovej komunikácie zahŕňa aj technológiu blockchainu, za účelom zlepšenia parametrov Device-to-Device komunikácie.

II. MOBILNÉ SIETE

A. Piatá generácia mobilných bezdrôtových sietí

Rýchly nárast mobilných zariadení, ktorý všetci vnímame, stojí za vývojom novej generácie mobilných bezdrôtových sietí piatej generácie. Veľký počet zariadení pripojených k mobilnej službe, zahľucuje základňovú stanicu a tá nedokáže obslužiť obrovský počet zariadení. Znižuje tým šírku pásma jednotlivých kanálov čo má za následok, že každé zariadenie má nižšiu prenosovú rýchlosť a vyššiu odozvu. Sieť piatej generácie sa budú deliť na dve samostatné siete pod 6 GHz a nad 6 GHz. Nízkofrekvenčné siete budú pracovať na frekvenciách pod 6 GHz a ponúknu jednotlivú šírku kanálov až 100 MHz. Druhá časť siete bude pracovať na frekvenciách nad 6 GHz, čím dôjde k použitiu nového spektra až nad 30 GHz, ktoré sa nazýva milimetrové vlny.

Frekvenčný rozsah siete piatej generácie je až do 100 GHz, pre Európu je sprístupnené frekvenčné pásmo od 24,25-27,5 GHz. V týchto pásmach je možné použiť šírku jednotlivého kanálu o veľkosti až 400 MHz, čím teoretická rýchlosť prenosu dosiahne hodnotu 20 Gbps.

Reálna prenosová rýchlosť však bude v rozmedzí 100 Mbps až 1 Gbps, ale záležať bude hlavne na veľkosti oblasti, hustoty zariadení a dosahu pokrytia prístupového bodu. Zväčšenie šírky kanálu a zvýšenie prenosovej rýchlosti spôsobí aj zníženie latencie na 1 milisekundu. Zlepšenie mobility a stabilného pripojenia až do rýchlosti 500 km/h a podpora obrovského počtu zariadení až $10^6/\text{km}^2$. Celkové zvýšenie spektrálnej účinnosti má byť oproti súčasným sieťam až 5x násobné [2] [3].

Prípady použitia 5G siete sa rozdeľujú do troch kategórií:

- Enhanced Mobile Broadband (eMBB) – Táto časť siete je venovaná mobilným zariadeniam. Dôležitým faktorom pri eMBB je vysoká prenosová rýchlosť, niekde medzi 10-20 Gbps. Dokážeme tak prenášať veľký obsah dát v krátkom čase, môžeme sledovať video v reálnom čase, môžeme hrať online hru s kamarátom vedľa nás taktiež v reálnom čase. Do tohto prípadu použitia spadá aj komunikácia Device-to-Device.
- Massive Machine Type Communication – Druhý prípad použitia umožňuje nasadenie obrovského počtu senzorov, kamier, robotov a mnoho ďalších zariadení, ktoré medzi sebou priamo komunikujú. Hlavnou požiadavkou je v tomto prípade podpora obrovskej hustoty zariadení až $10^6/\text{km}^2$ a prenosová rýchlosť do 100 kbps. Dáta, ktoré sa budú prenášať budú malé a nie je potrebné využívať vysoké rýchlosti prenosu. Tento prípad umožní nasadenie Machine-to-Machine komunikácie.
- Ultra-Reliability and Low Latency Communication – Tretí prípad použitia 5G siete zakladá na spoľahlivosti a rýchlosti doručenia dát. Dáta musia od zdroja do cieľa doraziť čo najrýchlejšie a musí sa zachovať ich integrita. Dôvodom je nasadenie komunikácie Vehicle-to-Vehicle. Kritická infraštruktúra, kde spolu priamo komunikujú vozidlá vyžaduje spoľahlivosť komunikácie a hlavne jej dostupnosť 99,99% času z dôvodu, že vozidlá medzi sebou potrebujú komunikovať v reálnom čase.

B. Priama komunikácia medzi zariadeniami

Najsľubnejšia nová technológia prenosu dát v sieťach piatej generácie je priama komunikácia medzi zariadeniami, Device-to-Device. Táto technológia bola vyvíjaná už v sieťach štvrtej generácie ale nenašla v nej uplatnenie. Jej použitie v nových sieťach však vylepšuje mobilné služby a zlepšuje kvalitu služieb. Device-to-Device komunikácia umožňuje zariadeniam komunikovať medzi sebou priamo bez prechodu komunikácie cez základňovú stanicu alebo prístupové body. Je takto možné prenášať dáta väčšou rýchlosťou a v kratšom čase. Prínosom Device-to-Device komunikácie je aj zvýšenie spektrálnej účinnosti základňovej stanice, cez ktorú komunikácia nemusí smerovať. Rozdelenie Device-to-Device komunikácie na základe pomoci základňovej stanice pri Device-to-Device komunikácii, je možné na D2D s pomocou základňovej stanice, s asistenciou základňovej stanice, s pomocou smerovača alebo na priamu D2D komunikáciu.

Z pohľadu spektra môžeme D2D komunikáciu rozdeliť na vnútropásmovú a mimopásmovú. Vnútropásmovú komunikáciu rozdeľujeme na Underlay alebo Overlay. Mimopásmovú komunikáciu sa rozdeľuje na kontrolovanú a autonómnu. Pri vnútropásmovej komunikácii, zariadenia medzi sebou komunikujú v licencovanom pásme spoločne s celulárnou komunikáciou. Vnútropásmová komunikácia má výhodu kontroly and celulárnym spektrom, čím sa zachováva vysoká kvalita služieb. Vnútropásmovú komunikáciu rozdeľujeme na Underlay a Overlay D2D komunikáciu [5].

Pri Underlay vnútropásmovej D2D komunikácii, komunikácia prebieha súčasne s celulárnou komunikáciou. Využívajú sa zdroje základňovej stanice a celá komunikácia je nastavená základňovou stanicou. Opätovné využitie rádiových zdrojov zabezpečí, že tieto zdroje nie sú premrhané a sú využívané naplno. Súbežný priebeh celulárnej komunikácie aj D2D komunikácie v tom istom časovom okamihu však spôsobí vzájomnú interferenciu medzi týmito dvomi typmi komunikácie [4].

Overlay vnútropásmová D2D komunikácia, taktiež prebieha v licencovanom spektre. Celulárna komunikácia a D2D komunikácia bežia v rozdielnom časovom okamihu. Vzájomná interferencia medzi komunikáciami nám v tomto prípade zaniká. Problém je že môže dôjsť k plýtvaniu drahocenných rádiových zdrojov, keďže budú priradené pre typ komunikácie nemusia byť vždy naplno využité [6].

Mimopásmová D2D komunikácia prebieha v nelicencovanom spektre. Takéto spektrum ma označenie ISM. Jedná sa o industriálne, vedecké a lekárske pásmo, kde môže komunikovať ktokoľvek. Toto pásmo využívajú napr. technológie ako Wi-Fi Direct, Bluetooth. Hlavnou nevýhodou použitia nelicencovaného pásma je jeho nevyspytateľná povaha. Keďže v pásme nie je žiadna kontrola od základňovej stanici dochádza tam k veľkej interferencii.

Mimopásmová kontrolovaná D2D komunikácia beží v nelicencovanom pásme, avšak nastavenie a čiastočná kontrola komunikácie má byť úlohou základňovej stanice, čím by mohlo dôjsť k zníženiu interferencie v nelicencovanom pásme. Pri mimopásmovej autonómnej D2D komunikácii sa celá komunikácia necháva na samotných zariadeniach. Nastavenie komunikácie, pridelenie rádiových zdrojov a riadenie komunikácie zodpovedajú samotné zariadenia. V nelicencovanom pásme je aj problém s objavovaním zariadení, a je potrebné použiť inú technológiu, ktorá pomôže vzájomne nájsť dve zariadenia, napr. Wi-Fi alebo Bluetooth [4] [5].

C. Technológia Blockchain

Blockchain je distribuovaná databáza, kde sa zaznamenávajú informácie o transakciách, ktoré sú organizované do hierarchického reťazca blokov. Prvá aplikácia, ktorá využíva technológiu Blockchainu je kryptomena Bitcoin. Využitie technológie blockchainu v iných odvetviach sa javí veľmi sľubne. Blockchain pozostáva z kryptografických, matematických algoritmov a ekonomického modelu. Celkovo sa skladá zo šiestich kľúčových elementov [7]:

- Decentralizovaná štruktúra
- Transparentnosť
- Autonómia
- Nemennosť
- Anonymita
- Open source

Štruktúra jedného bloku v Blockchaine sa skladá:

- Data
- Hash
- Časová známka
- Ďalšie informácie

III. NÁVRH SIMULÁCIE

Návrh simulácie viacpreskokovej D2D komunikácie je realizovaný v programovacom jazyku Python. Pre simuláciu bolo potrebné aj navrhnuť možnosti zvýšenia robustnosti komunikácie, cieľom bolo zlepšiť prenosovú rýchlosť, priepustnosť komunikačného kanála, znížiť stratovosť paketov. Najskôr boli stanovené rozmery oblasti, kde sa budú jednotlivé zariadenia nachádzať na 1000x1000m. V oblasti sa náhodne vygeneruje pozícia uzlov, ktorých počet je zadaný, napr. 100. Každý uzol má aj faktor dôveryhodnosti. Tento faktor určuje či je uzol spoľahlivý na posielanie dát. Vypočítava sa z prenosu úspešných paketov. Ak sú pakety úspešne poslané do ďalšieho uzla, nestratia sa, zachovávajú svoju integritu, tak sa tento faktor zvyšuje, naopak, ak sa pakety na danom uzle strácajú, sú pozmenené a tento faktor znižujú. Cesta medzi zdrojovým a cieľovým uzlom je vyberaná na základe tohto faktoru. Jedná sa o najlepšiu trasu, ktorá nie vždy je najkratšia. Po nájdení trasy sa medzi zdrojovým a cieľovým uzlom začnú posielat' pakety troch veľkosti, 100 kb, 200 kb and 300 kb. Merá sa čas potrebný k prenosu, prenosová rýchlosť, strata paketov.

Priepustnosť komunikačného kanála je počítaná na základe Shannon-Hartley teorému, kde kapacita kanála sa rovná šírke pásma vynásobenej logaritmom so základom čísla dva z pomeru signálu ku šumu zobrazené prostredníctvom nižšie uvedeného vzťahu. Kde C je kapacita kanálu v bitoch za sekundu, teoretická horná hranica čistej bitovej chybovosti, B reprezentuje šírku pásma, S popisuje priemerný výkon prijímaného signálu, N je priemerný výkon šumu a rušenia a S/N popisuje pomer signálu k šumu.

$$C = B * \log_2 \left(1 + \frac{S}{N} \right)$$

Navrhnutý model komunikácie bol testovaný prostredníctvom parametrov zobrazených v tabuľke 1.

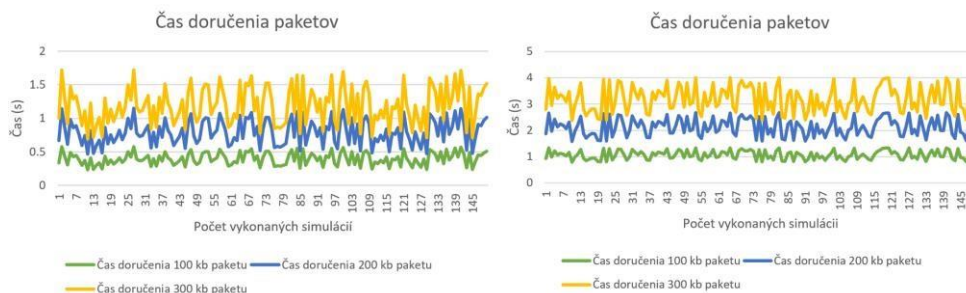
Tabuľka 1. Simulačné parametre

Premenná	Hodnota
Veľkosť oblasti	1000x1000 metrov
Počet uzlov	100
Šírka pásma	20 MHz
Pomer signálu k šumu	20-25 dB
Rádiový dosah	200 metrov
Faktor dôveryhodnosti	60%
Veľkosť paketov	100 kb, 200 kb, 300 kb

IV. VÝSLEDNÉ INFORMÁCIE SIMULÁCIÍ

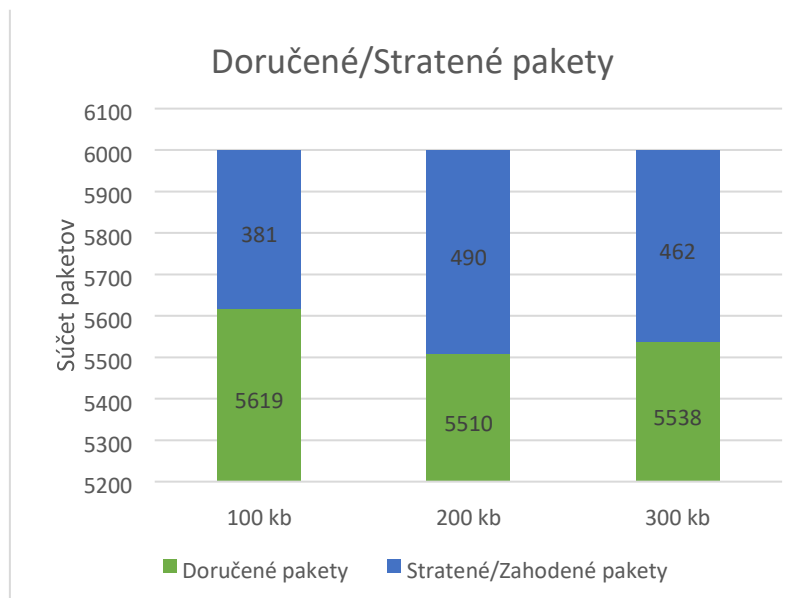
Celkový počet simulácií dohromady bol 600. Simulácie boli rozdelené do štyroch kategórii na základe počtu smerovačov medzi zdrojovým a cieľovým uzlom, takže v každej kategórii bolo vykonaných 150 simulácií. Prvá kategória mala medzi zdrojovým a cieľovým uzlom 1-2 smerovače, druhá kategória mala 3-4, tretia kategória 5-6 a štvrtá kategória 7 a viac.

Prvá kategória, ktorá mala na trase 1-2 smerovače je zobrazená na obrázku 1 vľavo. Na osi x sa nachádza počet vykonaných simulácií. Pre danú kategóriu to bolo 150 simulácií. Na osi y je zobrazený čas prenosu medzi zdrojovým a cieľovým uzlom. Zelenou farbou v grafe je čas prenosu 100 kb paketov, modrou farbou 200 kb a žltou farbou 300 kb. Z grafu vidno, že prenos väčšieho paketu trval dlhší čas. Priemerný čas doručenia 100 kb paketov bol 0,396 s, čas doručenia 200 kb paketu bol 0,793 s a priemerný čas doručenia 300 kb paketu bol 1,19 s.



Obrázok 1 Čas doručenia paketov, počet skokov 1-2 / 5-6

Podobný graf, zobrazený na obrázku 1 vpravo, však zobrazuje čas doručenia paketov medzi zdrojovým a cieľovým uzlom keď na trase bolo 5-6 smerovačov. Na pohľad je vidno, že väčší počet smerovačov na ceste spôsobil, že doručenie paketov trvalo dlhší čas. Priemerný čas doručenia 100 kb paketu bol 1,063 s, 200 kb paketu bol 2,126 s a 300 kb pakety trvalo doručenie v priemere 3,189 s. Celkové výsledky sú zobrazené na obrázku 2.



Obrázok 2 Graf doručených a stratených paketov

Prvý scenár s veľkosťou paketov 100 kb sa javí ako najlepší. Celkovo bolo prenesených 6000 100 kb paketov medzi zdrojovým a cieľovým uzlom. Bola dosiahnutá najmenšia stratovosť len 381 paketov bolo zahodených, pozmenených. Úspešne doručených bolo 5619 paketov, čo činilo 94.14% úspešnosť prenosu dát. Druhý scenár s veľkosťou paketu 200 kb je najmenej úspešný. Stratilo sa najviac paketov zo všetkých troch scenárov až 490, čo je len 92.16% úspešnosť prenosu. Pri treťom scenári posielania paketov o veľkosti 300 kb sa stratilo o niekoľko paketov menej ako v druhom scenári, 462 stratených paketov, percentuálna úspešnosť prenosu bola podobná 92.38% [1].

V. ZÁVER

Device-to-Device komunikácia je sľubnou technológiou v nových 5G sieťach. Výhody, ktoré so sebou prináša zlepšujú efektivitu mobilných služieb. Na základe rozdelenia je ju možné použiť nielen v súčasných celulárnych sieťach ale aj mimo nich, čo prispeje k väčšiemu dosahu signálu. Využitie D2D komunikácie v rôznej sfére spraví sieť dostupnejšie pre každého. S využitím technológie blockchainu je možné zlepšiť robustnosť siete. Teória spísaná v práci oboznamuje o nových bezdrôtových mobilných sieťach 5. generácie, rôznymi typmi mobilnej bezdrôtovej komunikácie, porovnanie s mobilnými ad hoc sieťami a technológiu blockchainu, potrebné pochopiť princípy k porozumeniu využitiu v návrhu modelu viacpreskokovej D2D komunikácii. Podľa tvrdení overených v simulácií bolo zistené, že je možné zlepšenie bežnej komunikácie cez technológiu D2D.

POĎAKOVANIE

Táto publikácia vznikla vďaka podpore v rámci Operačného programu Integrovaná infraštruktúra pre projekt: Výskum v sieti SANET a možnosti jej ďalšieho využitiu a rozvoja, kód ITMS: NFP313010W988, spolufinancovaný zo zdrojov Európskeho fondu regionálneho rozvoja.

POUŽITÁ LITERATÚRA

- [1] E. Šaršala. Diplomová práca. Model D2D komunikácie pre 5G siete, TUKE Košice, 2022.
- [2] Yu Tang, Sathian Dananjayan, Chaojun Hou, Qiwei Guo, Shaoming Luo, Yong He, A survey on the 5G network and its impact on agriculture: Challenges and opportunities, *Computers and Electronics in Agriculture*, Volume 180, 2021
- [3] A. N. Al-Quzweeni, A. Q. Lawey, T. E. H. Elgorashi and J. M. H. Elmirghani, "Optimized Energy Aware 5G Network Function Virtualization," in *IEEE Access*, vol. 7, pp. 44939- 44958, 2019
- [4] K. Venugopal, M. C. Valenti and R. W. Heath, "Device-to-Device Millimeter Wave Communications: Interference, Coverage, Rate, and Finite Topologies," in *IEEE Transactions on Wireless Communications*, vol. 15, no. 9, pp. 6175-6188, Sept. 2016
- [5] J. Guo, S. Durrani, X. Zhou and H. Yanikomeroglu, "Device-to-Device Communication Underlying a Finite Cellular Network Region," in *IEEE Transactions on Wireless Communications*, vol. 16, no. 1, pp. 332-347, Jan. 2017
- [6] Kim, DH., Oh, SJ. & Lim, J. Multi-channel-based scheduling for overlay inband device-to-device communications. *Wireless Netw* 23, 2587–2600 (2017)
- [7] Xu, M., Chen, X. & Kou, G. A systematic review of blockchain. *Financ Innov* 5, 27 (2019)

Online Voting Management System Based on BlockChain Technology

¹*Bianka BILECOVÁ*, ²*Ján HURTUK*

¹Department of Computers and Informatics, Faculty of Electrical Engineering and Informatics Technical University of Košice, Slovak Republic

¹bianka.bilecova@student.tuke.sk, ²jan.hurtuk@tuke.sk

Abstract – The topic of this paper is an online voting management system based on blockchain technology, which has been successfully designed and implemented. The theoretical part is focused on blockchain itself and online voting. The work results in a functional blockchain-based system that allows administrators to create new polls, start them and stop them when it's over. Voting can be public or secret. The system allows voter registration and subsequent voting. After the vote, a report is available depending on whether the vote was public or secret.

Keywords – blockchain, online voting, smart contracts, voting management system

I. INTRODUCTION

Online voting is a topic that the society has been dealing with for many years. Several countries has tried such type of voting and are still trying to implement it instead of traditional paper elections thanks to it's many benefits. These benefits include a reduction in the time and cost required to vote. The results of the voting are available almost immediately after the voting and the error rate in the counting of votes is also eliminated. Another big advantage is the increased participation of citizens, as online voting can overcome physical constraints, so all targeted voters can vote regardless of their actual location.

However, using central databases, online voting was dangerous and prone to attacks on the system. Blockchain introduced a new solution that was almost ideal for voting purposes. The use of blockchain has also increased the security and reliability of online voting systems. The voting became transparent.

Such voting does not have to be limited to national elections, it can be used for various purposes, such as various internal votings in companies and enterprises. The aim of this work is to create an online voting management system which will deal with such minor problems, whether it is a secret or a public vote. Blockchain provides a transparent voting solution, which ensures that the cast votes are not manipulable and are counted correctly.

II. ONLINE VOTING

Online voting is a type of voting in which a person can vote through an internet-connected device. It can be used in referendums, elections or for other purposes. It has a numerous advantages, such as a reduction in financial costs, the elimination of invalid votes or an increase in turnout and therefore such a method of voting has already been introduced in several countries around the world.

Despite its many advantages, it also has certain shortcomings and there are a number of requirements that the voting system must meet in order to be usable. Such a system should include voter registration, authentication, voting itself and counting of votes [1].

A. Requirements

There are a number of requirements that must be met in any electronic voting system. Wang and his team conducted a general and comprehensive survey, which resulted into dividing the online voting system requirements into the core and additional requirements [2]. They recommend to include all core requirements in the online voting systems and pick the other additional requirements depending on the voting environment.

Core requirements:

- Correctness (completeness and soundness)
- Privacy
- Unreusability
- Eligibility
- Robustness
- Verifiability
- Usability

Additional requirements include fairness, uncoerability, efficiency, mobility, vote-and-go, universal verifiability, e2e-verifiability and practicality.

III. BLOCKCHAIN

The distributed nature of blockchain avoids the single point of failure issues and offers high integrity for transactions and data against intentional and unintentional corruption. All users can use their devices to keep a large number of transaction records in blocks that construct the chain.

A. Fundamental principles of blockchain

Blockchain technology is based on certain principles which distinguish it from other technologies. According to Samar Al-Saqq and Sufyan Almajali, the blockchain is built on five fundamental key principles [3]:

- Decentralization
- Immutability
- Transparency
- Anonymity
- Consensus driven

Based on the main characteristics of the blockchain system, we can say that it is a suitable solution for use in voting systems. Hashing and concatenation of blocks in the network prevents data modification, thus ensuring the irreversibility of records. Because the central server is replaced by a group of independent nodes on the network, the problem of central point failure is also eliminated. It also provides voter transparency and anonymity, if necessary. In order for a block to be added to the network, there must be a consensus between the nodes. Thanks to smart contracts, clear rules can be set during the elections, which can prevent many conflicts.

B. Smart contracts

Smart contracts are digital, distributed contracts that run when predetermined conditions are met. They are small self-management programs stored inside the blockchain, which perform operations on the basis of predefined conditions between the parties [4]. These contracts serve as a transparent definition of the contract reducing the costs of verification, execution and fraud prevention. They took over the role of the centralized institutions.

The smart contract has a unique address, a set of executable functions and state variables [5]. The user starts a smart contract by addressing it a transaction. Subsequently, the contract is automatically and independently executed according to the predetermined instructions depending on the data contained in the current transaction.

IV. ANALYSIS OF BLOCKCHAIN PLATFORMS

We chose to compare three widely used blockchain platforms:

- Ethereum
- Hyperledger Fabric
- R3 Corda

After comparing these three platforms, Fabric and Ethereum stand on one side. Both are highly flexible, but in different respects. Ethereum's smart contracts make it a general platform for virtually any type of application. Because this platform is permissionless and completely transparent, there is poorer scalability and less privacy. Hyperledger Fabric solves scalability and privacy issues using the Byzantine Fault Tolerance algorithm and access control [6]. Corda was designed exclusively for the financial sector. This simplified its architectural design. There are efforts to integrate Corda into the Hyperledger Fabric so Corda cannot be considered a competitor but rather an adjunct.

After comparing those three platforms we chose to use Ethereum platform for our online voting system since this platform was most suitable for our needs.

A. Advantages of online voting systems based on blockchain

By using blockchain in online voting, we can make the entire election management more efficient [7]. We can use it for example to check whether voters have voted, counting votes and much more. We can also implement electronic administration, such as recording and storing voting rights in the blockchain.

Compared to conventional voting, a blockchain-based voting system can reduce the time and cost required to vote. There is no need to wait for the result of the vote, it is possible to see it almost immediately after the vote ends. As the voting process is simplified, we can also expect a reduction in voting costs [8]. Compared to non-blockchain-based online systems, the cost of implementing a central server and security system is reduced.

An online voting system can increase citizen participation in voting. For many people, it is difficult to take part in the current direct ballot. With the possibility of online voting, physical limitations can be overcome and more people can be involved in the voting process. Registered voters who own an electronic device can easily access the voting platform regardless of their actual position.

With regard to online voting systems, there are concerns about the secrecy of voting, issues of personal data security, abuse and falsification of the right to vote. However, the blockchain-based system does not use a central server as is the case with conventional voting systems. Use of a decentralized network has a lower risk of falsifying the election result. The vote is transparent.

V. DESIGN AND IMPLEMENTATION OF SOLUTION

A. System requirements

The system should support creation of new votes with a choice of several options. It should allow administrator to start, stop and monitor the progress of voting. System should be based on blockchain technology to make the voting process transparent. The administrator should have access to all votes, so he could verify the voting results. As voting will be based on blockchain, it will also be secure and votes stored on network will no longer be able to be altered or manipulated in any other way. The system will support both public and secret voting (individual votes will not have access to the voting results).

The proposed system should meet the following requirements:

- Support for user registration and login
- The ability to create new polls (public, secret) from the administrator's account
- The possibility to start and stop voting
- Summary of voting results

B. System architecture

Architecture of the system consists of frontend part which is React application and Spring boot application, there is also a database and ethereum blockchain. We can see how these parts interact on the next figure Fig. 1.

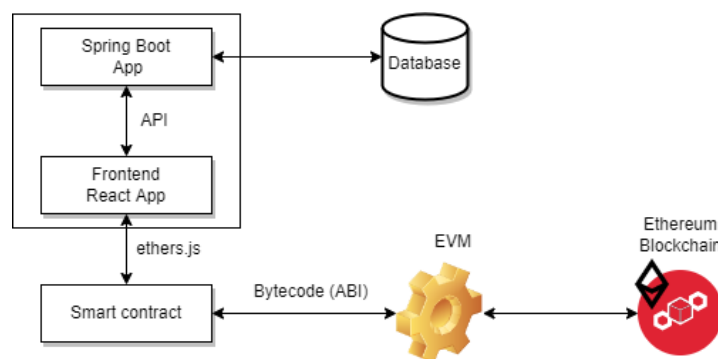


Fig. 1 Diagram showing the architecture of system interactions.

The Spring Boot application is used to communicate with the database, which stores the registration data of individual users. It also communicates with the frontend via the API and provides information about the success or failure of individual processes. Based on this communication, it is possible to register and log in users. Using the ethers.js library, it is possible to link to the smart contract based on its address in ethereum blockchain and ABI (application binary interface), which can be obtained after compiling the smart contract from the generated JSON file. The smart contract further communicates with EVM and subsequently with the blockchain itself.

C. User interface

To implement the user interface of the client part of our solution, we decided to use ReactJS which is currently one of the most popular JS libraries. Even after years of its release, it is consistently used by web developers. It simplifies the process of creating interactive UIs.

For the development of graphical user interface was used the Bootstrap framework, which sped up the process and made the graphic part of this implementation look nice and smooth. Created web application allows registration, login, creation of votes, voting itself and an overview which will be available at the end of the vote.

On the next figure, Fig. 2 we can see example of all votes screen from the view of the administrator. Here, all created votes are displayed together alongside with their current status, meaning whether the vote is created, in progress or already finished. Through this screen, the administrator can start and end the vote. There is also an option to display the details of individual votes.

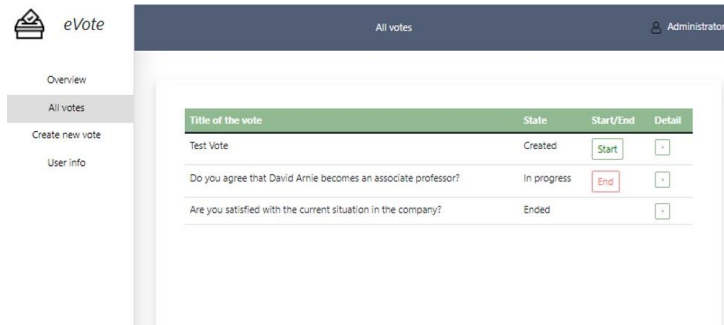


Fig. 2 Example of all votes screen.

D. Smart Contracts

For this solution were created two smart contracts. They were written in Solidity programming language. One of the smart contracts is for public vote and the other is for secret vote. These contracts are similar, differing only in some small details such as private variables or user IDs which are available only in a public vote. The smart contracts were deployed on the blockchain by the administrator and subsequently connected to the web application to allow mutual communication.

E. Truffle

The truffle framework is used in this implementation to test smart contracts and deploy them to the blockchain. The Truffle framework facilitates process of development, testing, and deployment of decentralized applications. It provides a development environment for the blockchain network. It can be used to create smart contracts, compile them, link and save them to the system. Ganache is part of the Truffle ecosystem.

F. Ganache

Since working with the main ethereum network costs actual money for transactions, we chose to use local RPC blockchain called Ganache. This local test network is suitable for the entire development cycle. It enables us to develop, deploy and test our decentralised application in a safe and deterministic environment. We chose to use ten test accounts each having 100 ethers for testing purposes.

Ganache comes with two components to make it user-friendly, Ganache CLI and Ganache GUI. Ganache CLI allows us to create our own local blockchain environment using terminal commands. Ganache GUI can be generally described as a blockchain explorer.

G. Spring Boot application and database

Users can register using their name, email and password. This data will then be stored in a database and will be used for authentication when individuals log in. The password is stored in the form of a SHA-256 hash. Communication with the database takes place via the Spring Boot application. During registration, it is verified whether the given user already exists and if so, the web application will receive a response with the message that the user registered to such an email already exists.

H. React application and blockchain

React application is designed to communicate with the smart contracts. Inside a smart contract there are created events. Events in smart contracts are mainly used to communicate with decentralized applications and other smart contracts in case there is an important change they should be informed about [9]. In this solution, on the side of the smart contract, an event is defined, which is emitted when a specific function is called and on the decentralized application side, there is a listener for a specific event.

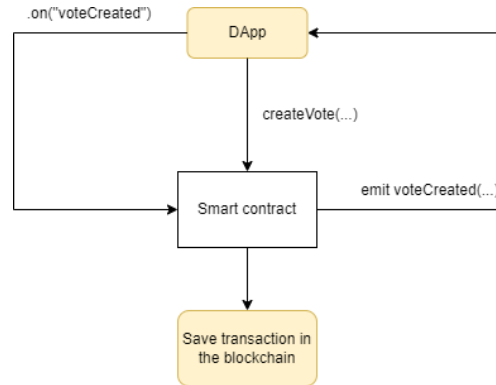


Fig. 3 Diagram showing the process of updating UI.

Fig. 3 shows how the user interface updates works. These updates take place after a vote has been created, a vote has been cast and after every start and end of the vote.

I. Public vote

In a public vote, the data of each user who cast its vote is also available to the administrator along with the individual's vote. This is accomplished by linking user's id and address. When casting a vote, the user also sends its id, which will then be returned via the event. The administrator can associate this id with the name in the database to obtain more detailed voting information.

VI. CONCLUSION

The system has gone through and passed all the test scenarios, so we can say that the system is functional and met all the requirements placed on its development. Of course system still has room for optimization and improvement. In the future, better securing of secret votes would be appropriate, for example by implementing some of the cryptographic functions. And as far as registration process goes, better security of the system could be achieved by approving individual users by central authorities before giving them access to the system.

REFERENCES

- [1] U. Jafar and M. Aziz, "Blockchain for electronic voting system—review and open research challenges," *Sensors*, vol. 21, no. 17, 2021.
- [2] K. Wang, S. Mondal, K. Chan, and X. Xie, "A review of contemporary evoting: Requirements, technology, systems and usability," *Data Science and Pattern Recognition*, vol. 1, no. 1, pp. 32–34, 2017.
- [3] S. Al-Saqqa and S. Almajali, "Blockchain technology consensus algorithms and applications: A survey," *International Journal of Interactive Mobile Technologies*, vol. 14, no. 15, pp. 142–156, 2020.
- [4] J. Strebko and A. Romanovs, "The advantages and disadvantages of the blockchain technology," *6th Workshop on Advances in Information, Electronic and Electrical Engineering*, pp. 1–6, 2018.
- [5] M. Gates, *Blockchain: Ultimate guide to understanding blockchain, bitcoin, cryptocurrencies, smart contracts and the future of money*. CreateSpace, 2017.
- [6] M. Zand, X. Wu, and M. Morris, *Hands-On Smart Contract Development with Hyperledger Fabric V2: Building Enterprise Blockchain Applications*. O'Reilly Media, 2021.
- [7] R. Kim, K. Min, and S. Hong, "A study on ways to apply the blockchain-based online voting system," *International Journal of Control and Automation*, vol. 10, no. 12, pp. 121–130, 2017.
- [8] V. Gatteschi, F. Lamberti, C. Demartini, C. Pranteda, and V. Santamaria, "To blockchain or not to blockchain: That is the question," *IT Professional*, vol. 20, pp. 62–74, 2018.
- [9] G. Oliva, A. Hassan, and A. Jiang, "An exploratory study of smart contracts in the ethereum blockchain platform," *Empirical Software Engineering*, vol. 25, pp. 1864–1890, 2020.

Design optimized flow collector for network capturing

¹Marek GLEVAŇÁK, ²Martin HASIN, ²Martin CHOVANEC, ²Martin HAVRILLA

¹Department of Computers and Informatics, Faculty of Electrical Engineering and Informatics Technical University of Košice, Slovak Republic

²Department of Computers and Informatics, Faculty of Electrical Engineering and Informatics Technical University of Košice, Slovak Republic

³Department of Computers and Informatics, Faculty of Electrical Engineering and Informatics Technical University of Košice, Slovak Republic

⁴Department of Computers and Informatics, Faculty of Electrical Engineering and Informatics Technical University of Košice, Slovak Republic

¹marek.glevanak@student.tuke.sk, ²martin.hasin@tuke.sk, ³martin.chovanec@tuke.sk, ⁴martin.havrilla@tuke.sk

Abstract – In the past years network was spreading. The network infrastructures are complicated and the risks of cyber attacks are highly possible. With a growing number of devices, network traffic and bandwidth of this traffic are also spreading. Main task of this thesis is to try to design dropless flow collector from incoming network traffic in our infrastructure. We will try to achieve that with a technique of scaling Linux network stack and also with customizing the parameters of network card. We will talk about DPI which will be used to get the information from incoming traffic in our collector. This technique will give us not only simple information about traffic but also information about higher layers and protocols used in network infrastructure. This information will later be sent to elastic-search database. We will test designed flow collector on generated traffic. This traffic will be generated by TRex traffic generator at 10Gb/s speed rate. Outcome of this thesis will be flow collector which will be running on commodity hardware and also will be giving us statistical analysis about incoming traffic. We will test this in real network infrastructure.

Keywords – DPI, flow collector, network optimization, packet generators

I. INTRODUCTION

In the past 25 years network has changed world. We live in the times when we cannot imagine world without network. In these years network is used by almost 5 billion people. This huge amount of user also means many potential victims of cyber crimes. With trends like "BYOD" or bring your own device, many of users are using their own devices in companies or school network infrastructures. This opens opportunities for attacker to easier penetrate through the secure infrastructure and attack them from inside [1], [2].

With these facts we decided to implement device to monitor traffic. This device is called the flow collector and it will harvest data from incoming packets and provide us statistical information about traffic flowing through our infrastructure. But with a growing network and many users connected in the infrastructure, the bandwidth of network inside this infrastructure is high. A specialized hardware for monitoring traffic and providing such functionality cost a lot of money. We want to move to the commodity hardware to provide specialized function as was presented.

The task of this work is to try to design a lossless flow collector. We will consider lossless as a goal that we would like to achieve. Although we know this goal is very challenging, we would like to get closer to it. Our flow collector will be used to obtain information from the data stream which we will then be able to use for looking for anomalies in the data flow and also to have factual information in the event of an incident in our network. With high bandwidth we need to ensure lossless state so we will be working on techniques to achieve high throughput in our flow collector and we will present the technique to scale flow collector to achieve dropless state. We will design this flow collector also in such manner and present a technique to analyze the data from flowing network.

II. NETWORK TRAFFIC ANALYSIS

Nowadays, many tools for extracting data and metadata from network traffic exist. Data classification from incoming traffic is a hard task, even though we have very fast computing device. This task is even harder when we want to detect closely from packets, what protocols are there in traffic. DPI have a many usecases for extracting information from flowing traffic, but the biggest implementation what we can see, is in China. All incoming traffic is inspected and if data are censured, traffic will be blocked by this DPI [3].

Older tools they analyzed such a flow in a simple way by looking at the destination port in the packet header, and determining the protocol accordingly. If we want to use an example, with TCP connection and port 22 we can say that it is an ssh protocol. However, the accuracy of this technique has began to fall dramatically as protocols, that dynamically change ports, increased. Another feature that has reduced the quality of this technique is port sharing and redirection. This brought us to the point, where we need to know something more about packet. We need to know more about the data itself which are hidden in the packets [4]. Such technique is beneficial for us, not only because we know how to identify what flow passes through our facilities, but mainly for security in our network [5]. Deep packet inspection is a technique that has gained great popularity today. It provides the ability to classify and control packet flow in the terms of application content [6].

Algorithms that look for a string of characters have proven to be very beneficial as in the network context, as well as in DPI. This technique requires scanning each byte of the data contained in the packet, and then look in predefined string or a pattern. Such a process is, of course, it was very demanding, so several algorithms were devised to search for patterns in data such as Aho-Corasick [6], [7].

A. Handling secure traffic

TLS is the most popular technique for encrypting data over the network, and of course also the most used. TLS replaced its predecessor SSL. The problem occurs with the gradually increasing trend of sending data over encrypted communication. This trend is not bad at all, but it makes difficulties to detect a malicious stream of data. Like the network providers, as well as cloud providers, one of their primary objectives is to provide the service in a safe way. At this point, the techniques of DPI itself prove weak, and the search for strings in the encrypted packet stream is completely worthless.

A number of corporate companies are using TLS proxy [8] to be able to own TLS certificates to provide security [9] and at the same time be able to see the whole stream unencrypted. This can help to look for anomalies or potential attacks in it. Unfortunately, when potentially attacked servers work as a TLS proxy, an attacker is able to see the unencrypted flow, which can have a huge impact on employees and the company itself.

The TLS cannot be decrypted, but we can use TLS protocol to help us find out more about traffic itself. One of the common misconception is that if we use TLS, the server name is also encrypted. Yet, the monitoring applications such as Wireshark or network providers, which can intercept the network traffic, can still find out the server name that we are asking for. This server name information is sent during the initialization packet by the client and the specific value is called SNI [10].

B. nDPI

Commercial paid instruments that provide DPI are very expensive, whether in terms of maintenance or license value. Since the commercial version do not offer open source, it is impossible for a developer or customer to add their own protocol to this tool. Also public DPI tools are mostly just limited versions that have been developed to validate research theory. nDPI is an open-source library that provides DPI. It also aims to use this DPI on commodity hardware, because most researches aim to use a specialized hardware or specialized network card with FPGA. With nonspecialized hardware we can achieve analyzing traffic with up to 10Gb/s. nDPI can also handle encrypted traffic by implemented decoder which can extract server name from server certificate issued by SSL [5].

C. NFStream

NFStream is an open-source framework that use underhood nDPI. This allow to have high throughput network traffic flow analysis to be run on commodity hardware [11]. NFStream is created in python to be more flexible and more user friendly with faster learning curve. Incoming packets are aggregate to flow by shared parameters for each packet, for example source and

destination port, source and destination IP. From this flow it can provide statistical information later.

NFStream can operate offline from pcap file and also online from network card. Both operations use libpcap as standard capture device. High speed in online mode is achieved by parallelism with modified version of libpcap. This modified version is using AF_PACKETv3 with technique called fan-out, which can distribute load among all threads which are using fan-out socket [12], [11].

PACKET GENERATORS

The test flow on which we will test the flow and processing of packets is important part of the proposal.

As the popularity of the network among people grows, so do many the network services high demands in terms of security, scalability and availability. Packet generators have also found their place in testing in this area. The main task of the generator packet is to allow the desired, controlled packet flow and thus simulate a task on this network. One of the disadvantages of packet generators is that the source file has to be captured in real network infrastructure [13].

Packet generators are implemented in both software and hardware solutions. The disadvantage of hardware packet generators is, of course, their price [14]. With today's hardware equipment we can run software solution generator and scale it on the necessary activity. So the packet generators know how to emulate different parts of network layers. From this we can conclude that these generators can also replace full-fledged traffic such as HTTP, HTTPS and various other higher layer protocols.

While high-speed packet generators can send simple data from predefined pcap files, more advanced generators can send more complex network proposals and can respond to them as well. Mostly it is implementing via the API defined by such a complex network generator. Various researchers use packet generators to test their hypotheses as proof-of-concept [14], [15].

TRex packet generator

TRex is an open-source project. This project focuses on generating an network traffic that is suitable for stress testing network and is also suitable for testing devices on this network.

This generator can emulate from L3 to L7 layer, which means that we can test different cases in this network and thus prepare the network or network devices to production environment. An example is the firewall testing when the network is loaded at 10Gb/s or we can also simulate a TCP SYN attack to server. Last but not least, the advantage of this generator is that it can emulate higher network layers which will help us in researching and testing ours flow collector running on the server [16].

III. TOPOLOGY DESIGN FOR NETWORK RECORDING

The flow could be described as an undefined sequence of packets that passes through network devices such as routers, switches, and more. However, this undefined sequence passing through these devices has common features. We can use these features to obtain the necessary information about network communication in our devices, such as what type of data passing through the device or what type of data passes through our network [17].

Using the data collected from the flow collector, we can determine when to increase or decrease throughput in our network in the future. We can also use such data to detect anomalies in our network [17] [18] and prevent various problems whether with a poorly configured switch or an external attack such as DDoS attack, and also capture and record port scans.

In order not to unnecessarily increase the latency of the network we decided to place the collector outside the network which we will be monitoring. Another thing that needs to be defined is the output. With this experiment we would like to achieve more than we can with an ordinary flow collector. We would like to have more metadata on the output than we can achieve by just analyzing the packets. One of the options we have been considering is the packet analysis using deep and machine learning techniques. After a thorough analysis we found that such models are more accurate on the data we have already captured and not on the data we need to analyze over time. Another disadvantage of this a model is that we have to train it well on a large sample of data and even that does not guarantee the accuracy of the result [5].

As we mentioned, we want to use commodity hardware for purpose of collecting flow from traffic. This traffic will be analyzed by NFStream and than result from this analysis will be send to elasticsearch database. Hardware specification is described in table 1.

The main design decision was to use NFStream. This decision was made because it can provide us fast packet processing and friendly user space. In Fig. 1 we can see component design of our flow collector.

Table 1: Technical specification of server (flow collector)

RAM: 64GB
CPU: Intel(R) Xeon(R) CPU E5-4610 0 @ 2.40GHz, 48 core, 4 SOCKETS
Kernel version: 5.4.0-104-generic
Network card: Intel ® Ethernet Converged Network Adapter X520 10Gb/s, dual-port

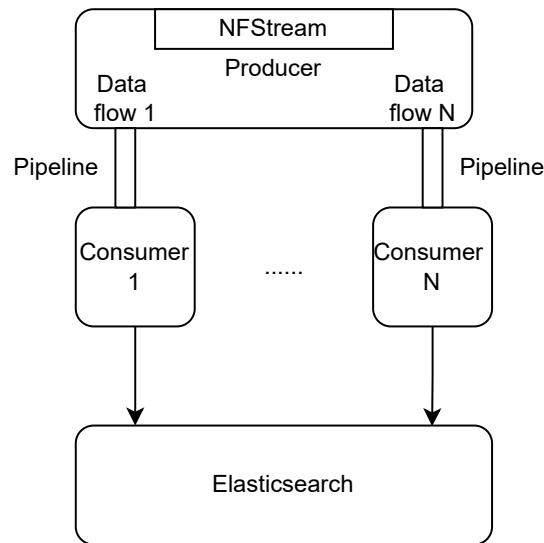


Fig. 1: Component design of flow collector.

Producer

The main responsibility of this component will be getting informations from NFStream, then filter unnecessarily data and pass the data into to the pipeline. Producer is lightweight wrapper around NFStream library with function which will be producing data forever.

Consumer

Consumer will be running on separate thread due to performance issues. We tested producer module as a standard API interface which was running on same thread as producer, but we got low performance. Consumer will consume data from pipeline module and sending this data to elasticsearch database. Consumer also has his own elasticsearch database module which provide caching for data. This cache then flush data after N items via `bulk_api` which is provided by elasticsearch library.

Pipeline

As a pipeline module we inherited queue from project `faster-fifo` [19]. We built wrapper around this which will provide set and get API to set flow, and get flow data. We have also created the `PipelineManager` which creates separate pipeline for every Consumer. On `set_flow` function call provides `round_robin` algorithm to store data evenly across consumers.

OPTIMIZATION OF NETWORK ELEMENTS

For the fast packet capturing we need to provide some specialized hardware. We chose to use fast network card from Intel company which can scale input processing up to 10Gb/s. Because we use specialized network card, we need to install driver for this card. We use `ixgbe` driver with the version 5.13.4. To scale up to 10Gb/s we also need to know more information on how to achieve this goal, because Linux comes with default parameters.

LRO and GRO

The problem with 10Gb/s Ethernet is that despite the speed between the lines, the transmission length limit is a maximum of 1500 bytes. This length can be circumvented by using so-called jumbo frame mechanism that can process up to 9000 bytes frame. For local networks where we can influence everyone to communicate with this size, it is very good option. For end users, where most are using predefined settings which is 1500 bytes frame, the jumbo frame setting can even cause a loss and performance degradation. So if we don't use a larger maximum length, we may pretend to use a bigger length [20] [21]. One of this technique is called LRO or large receive offload. LRO aggregates multiple packets into the large one and this bigger packet is then processed by operating system. But LRO is not sufficient enough because it may aggregate

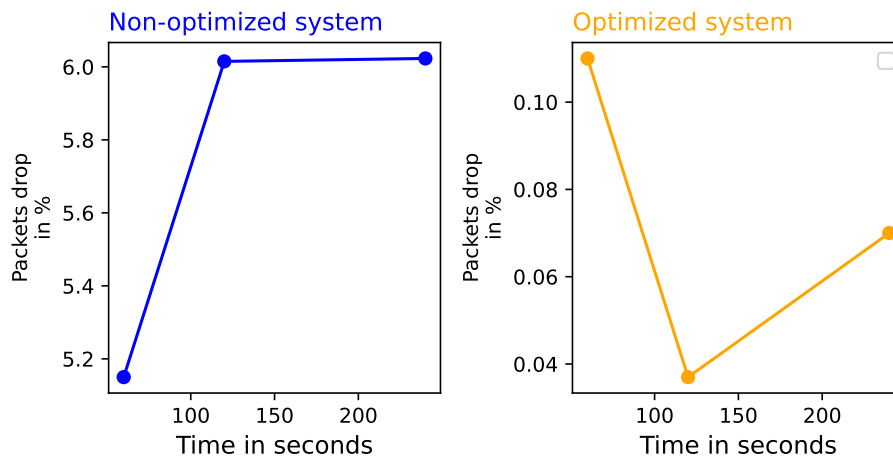


Fig. 2: Graph of packets drop in test environment with optimized and non-optimized system.

packets which have no connection. Solution for this problem is generic receive offload (GRO). GRO has a strict rules to aggregate incoming packets. It is also not limited only to TCP and IPv4 like LRO [20].

RSS

Receive Side Scaling (RSS) is a technique which support receiving data on multiple hardware queues from network card. If a network card with RSS enabled detects an arrival packets, it applies a filter to the packets and divides them into receive lines. The filter is a hash function which these packets pass through [22] [23]. With CPU affinity we can control from which queue or queues can which CPU process packet. Such optimization have big impact in NUMA architectures which are usually used in servers.

RPS is a software implementation of RSS in the core of the operating system. It means that we can turn this support with any network adapter that provides only one piece of hardware queue. This method does not make easier packet processing for example with NAPI, but it can quickly distribute packets to the defined processor.

Scaling of server

Optimizing network resources on our server is extremely important. We need to apply techniques we discussed above, to scale out server to get to maximum throughput from inbound traffic. First we verified the speed between the nodes on the basic settings that reached 8.90Gb/s, but with the fact that software interruptions reached 98% on one thread. We turn on GRO by command `ethtool -K ens5f0 gro on`. We tested this throughput with `iperf3` command and we achieved speed from **8.90Gb/s** at **98%** of CPU to **9.37Gb/s** and **0.7%**.

Next we distribute load by RSS technique. We applied RSS to our server and set one hardware queue to one core. In our case it was 48 hardware queues. We did this with command `ethtool -L ens5f0 rx 48`. The last thing we did was maximizing one hardware queue to maximum number. We did this by command `ethtool -G ens5f0 rx 4096` which increased size of one hardware queue from 512 to 4096.

TESTING FLOW COLLECTOR

With implemented flow collector we started testing our design. We applied all techniques mentioned above to achieve maximum process of input traffic. We also set NFStream to use maximum cores to gain maximum speed of receiving data.

In first experiment we decided to use smaller *active_timeout* and bigger *idle_timeout*. This variable was changing due to small test time and repeated traffic pattern, because we wanted to test packet drop under lot of pressure in short period of time. We know that generator generates very homogeneous traffic and for our purpose it does not make sense to do longer tests.

This test was simulated in our infrastructure with TRex packet generator and took 240 seconds under 10Gb/s. We tried to create generous traffic with protocol as VoIP, HTTPS, HTTP, CITRIX and many others. On the Fig. 2 we can see a graph of result from this tests. With our flow collector we achieved good behaviour of our software. We could see that we started from drop around 0.10% than drop started to decrease to 0.04% and with time drop slightly increased to 0.07%.

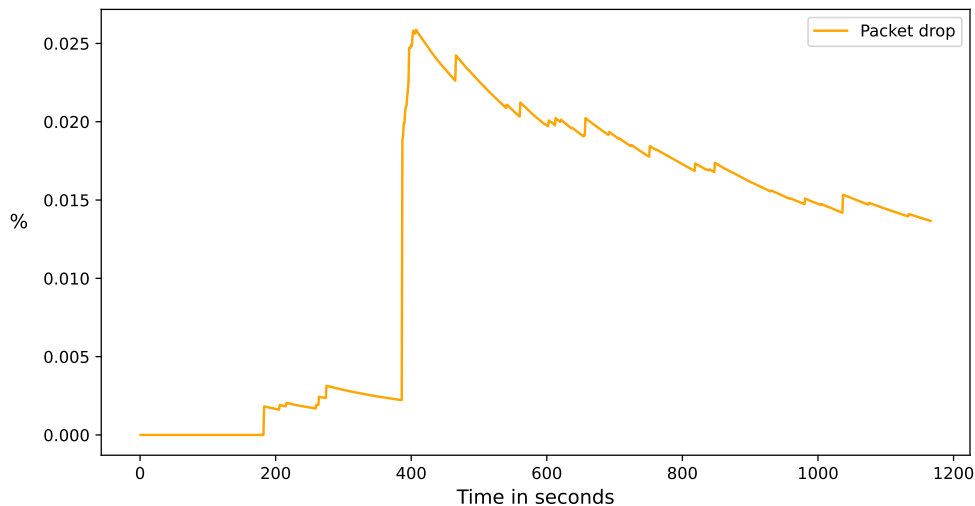


Fig. 3: Graph of drop packet in test environment.

We also compared optimized system to non-optimized system. To test non-optimized system we simply decrease number of hardware queues from 4096 to 512 and we turn off GRO. With such small changes we can see big difference in our test environment. We achieved drop around 6% which we can say is not reliable software. Non-optimized graph is on right side on Fig. 2.

After we finished testing with simulated environment we moved to real infrastructure. We use technique called port mirroring which mirror all data from one port to another without any change to core switch. But first we need to change values which was changed in testing environment. At the first place we changed *active_timeout* to 300 seconds and *idle_timeout* to 120 seconds. The longest flow will be captured at maximum 300 seconds. The core infrastructure give us from 200000 to 500000 packet per second and the result from this experiment can be seen on Fig. 3. As we can see we got better results compared to the test environment. This test took 1200 second and the highest achieved drop was 0.025%. Then drop started to decrease to 0.015% with occasional jump.

IV. CONCLUSION

The aim of this paper was to create a system that can capture data from incoming traffic and generate statistical data from them. This data will then be stored in a database with easy access. As a result we had to find out how DPI works and what are its advantages and disadvantages. We also analyzed the NFStream software, on which we built our software and which helped us from DPI integration. We described techniques how to achieve network optimization and we proved this with tests. In our test environment we showed there is a big difference between an optimized and a non-optimized system.

The final state of our work was to test the system in our infrastructure. This test lasted approximately 20 minutes and we have achieved a satisfactory result, even though this number is affected by various parts, which is difficult to rule out such as collisions and CRC errors. From this results we could say that the collector has proven itself in practice and achieves favorable results.

To achieve dropless we need to integrate kernel bypass technique to our software. This technique is designed to help fast packet processing in commodity hardware. This may help us capture all packets from wire and analyzed them. With such integration we could achieve lossless also in high loaded network environment.

ACKNOWLEDGMENT

This publication has been published with the support of the Operational Program Integrated Infrastructure within project: Research in the SANET Network and Possibilities of Its Further Use and Development (ITMS code: 313011W988), co-financed by the ERDF. Development and innovation of TUKE simulation environment in the field of technical sciences (004TUKE-2-1 / 2021)

REFERENCES

- [1] Y. Wang, J. Wei, and K. Vangury, "Bring your own device security issues and challenges," in *2014 IEEE 11th Consumer Communications and Networking Conference (CCNC)*, 2014, pp. 80–85.

- [2] W. Kim, O.-R. Jeong, C. Kim, and J. So, "The dark side of the internet: Attacks, costs and responses," *Information Systems*, vol. 36, no. 3, pp. 675–705, 2011, special Issue on WISE 2009 - Web Information Systems Engineering. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0306437910001328>
- [3] F. Yang, "The tale of deep packet inspection in china: Mind the gap," in *2015 3rd International Conference on Information and Communication Technology (ICoICT)*, 2015, pp. 348–351.
- [4] T. Bujlow, V. Carela-Español, and P. Barlet-Ros, "Comparison of deep packet inspection (dpi) tools for traffic classification," 2013.
- [5] L. Deri, M. Martinelli, T. Bujlow, and A. Cardigliano, "ndpi: Open-source high-speed deep packet inspection," in *2014 International Wireless Communications and Mobile Computing Conference (IWCMC)*, 2014, pp. 617–622.
- [6] S. Kumar, S. Dharmapurikar, F. Yu, P. Crowley, and J. Turner, "Algorithms to accelerate multiple regular expressions matching for deep packet inspection," *SIGCOMM Comput. Commun. Rev.*, vol. 36, no. 4, p. 339–350, Aug. 2006. [Online]. Available: <https://doi.org/10.1145/1151659.1159952>
- [7] P.-C. Lin, Y.-D. Lin, Y.-C. Lai, and T.-H. Lee, "Using string matching for deep packet inspection," *Computer*, vol. 41, no. 4, pp. 23–28, 2008.
- [8] L. Waked, M. Mannan, and A. Youssef, "The sorry state of tls security in enterprise interception appliances," *Digital Threats*, vol. 1, no. 2, may 2020. [Online]. Available: <https://doi.org/10.1145/3372802>
- [9] M. O'Neill, S. Ruoti, K. Seamons, and D. Zappala, "Tls inspection: How often and who cares?" *IEEE Internet Computing*, vol. 21, no. 3, pp. 22–29, 2017.
- [10] L. Deri and D. Sartiano, "Using dpi and statistical analysis in encrypted network traffic monitoring," *International Journal for Information Security Research*, vol. 10, pp. 932–943, 12 2020.
- [11] Z. Aouini and A. Pekar, "Nfstream: A flexible network data analysis framework," *Computer Networks*, p. 108719, 2022.
- [12] N. Bonelli, S. Giordano, and G. Procissi, "Enabling packet fan-out in the libpcap library for parallel traffic processing," in *2017 Network Traffic Measurement and Analysis Conference (TMA)*, 2017, pp. 1–9.
- [13] S. Molnár, P. Megyesi, and G. Szabó, "How to validate traffic generators?" in *2013 IEEE International Conference on Communications Workshops (ICC)*, 2013, pp. 1340–1344.
- [14] P. Emmerich, S. Gallenmüller, G. Antichi, A. W. Moore, and G. Carle, "Mind the gap - a comparison of software packet generators," in *2017 ACM/IEEE Symposium on Architectures for Networking and Communications Systems (ANCS)*, 2017, pp. 191–203.
- [15] P. Emmerich, S. Gallenmüller, D. Raumer, F. Wohlfart, and G. Carle, "Moongen: A scriptable high-speed packet generator," in *Proceedings of the 2015 Internet Measurement Conference*, ser. IMC '15. New York, NY, USA: Association for Computing Machinery, 2015, p. 275–287. [Online]. Available: <https://doi.org/10.1145/2815675.2815692>
- [16] team Trex. (2021) Trex realistic traffic generator. [Online]. Available: <https://trex-tgn.cisco.com/>
- [17] E. B. Claise, "Cisco Systems NetFlow Services Export Version 9," Internet Requests for Comments, RFC Editor, RFC 1654, October 2004. [Online]. Available: <https://www.rfc-editor.org/rfc/rfc3954.txt>
- [18] E. Garšva, N. Paulauskas, G. Gražulevičius, and L. Gulbinovič, "Academic computer network traffic statistical analysis," in *2012 2nd Baltic Congress on Future Internet Communications*, 2012, pp. 100–105.
- [19] A. Petrenko and T. Kumar, "A faster alternative to python's multiprocessing.queue," <https://github.com/alex-petrenko/faster-fifo>, 2020.
- [20] C. Jonathan. (2009) Jls2009: Generic receive offload. [Online]. Available: <https://lwn.net/Articles/358910/>
- [21] L. Jie, C. Shuhui, and S. Jinshu, "Implementation of tcp large receive offload on multi-core npu platform," in *2016 International Conference on Information and Communication Technology Convergence (ICTC)*, 2016, pp. 258–263.
- [22] Ramneek, S.-J. Cha, S. H. Jeon, Y. J. Jeong, J. M. Kim, and S. Jung, "Analysis of linux kernel packet processing on manycore systems," in *TENCON 2018 - 2018 IEEE Region 10 Conference*, 2018, pp. 2276–2280.
- [23] D. Joe. (2016) Monitoring and tuning the linux networking stack: Receiving data. [Online]. Available: <https://blog.packagecloud.io/eng/2016/06/22/monitoring-tuning-linux-networking-stack-receiving-data/>

Automatická detekcia miest relevantných pre zmenu rečníka v dialógových interakciách

Stanislav ONDÁŠ, Silvia BAČIKOVÁ

Katedra elektroniky a multimedialných telekomunikácií, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovensko

¹stanislav.ondas@tuke.sk, ²silvia.bacikova@student.tuke.sk

Abstrakt — Miesto relevantné pre zmenu rečníka (TRP) predstavuje miesto v rozhovore, na ktorom môže nastať zmena rečníka. Výskyt a využitie týchto bodov v dialógu zabezpečuje zachovávanie plynulého striedania medzi rečníkmi. V predloženej práci sme sa zamerali na štúdium prozodických parametrov prehovorov v slovenskom jazyku a experimentálne sme overili potenciál týchto parametrov detegovať TRP. V prostredí Matlab sme porovnali rôzne klasifikátory na báze strojového učenia v úlohe automatického detektora TRP na báze výšky tónu a intenzity. Ako vyplýva z dosiahnutých výsledkov, prozodické parametre umožňujú detekovať TRP s akceptovateľnou presnosťou, ak boli trénované pre špecifického rečníka.

Príslušné slová — analýza konverzácie, striedanie iniciatívy, miesto relevantné pre zmenu rečníka, TRP, prozodické vlastnosti, klasifikácia

I. ÚVOD

Množstvo ľudských sociálnych aktivít si vyžaduje používanie pravidiel pre striedanie iniciatívy (turn-taking) určujúce poradie, v ktorom sa majú vykonávať jednotlivé akcie a kým sa majú vykonávať. Keďže je náročné hovoriť a počúvať súčasne, rečníci v dialógu musia nejakým spôsobom koordinovať to, kto v danom momente hovorí a kto počúva. Aj keď sa konverzačné systémy v rôznych formách stávajú čoraz viac bežnými, je zrejmé, že striedanie iniciatívy v týchto systémoch nie je ešte stále úplne zvládnuté. Tieto systémy majú často tendenciu nesprávne prerušovať používateľa alebo majú veľmi dlhé oneskorenie v reakcii. Modelovanie striedania rečníkov počas dialógových interakcií je teda stále do značnej miery predmetom aktívneho výskumu [1].

Dialógové výmeny (turns) a ich striedanie medzi rečníkmi sú základnými stavebnými kameňmi sociálnej interakcie. Koordinácia striedania rečníkov sa spolieha na osobnú analýzu aktuálneho rečníka spoluúčastníkom, teda aktuálnym poslucháčom, a na identifikáciu možného ukončenia prehovoru rečníka na základe syntaktických, pragmatických a prozodických podnetov. Načasovanie dialógovej výmeny v konverzácií možno považovať ako extrémne rýchle vzhľadom na kognitívne požiadavky na rečníkov a to pochopiť, naplánovať a vykonať obrat v reálnom čase [2].

Zaužívaným pojmom v rámci analýzy konverzácie je taktiež konštrukčná jednotka dialógových výmen (Turn-Constructional Unit, TCU). Ide o základnú stavebnú jednotku, z ktorých sú vyskladané jednotlivé dialógové výmeny medzi rečníkmi. Príkladom TCU sú veta, fráza, súvetie alebo jednoduchá jednoslovná odpoveď. Koniec akejkoľvek TCU predstavuje bod, v ktorom môže ďalší rečník prejavíť záujem o prebratie iniciatívy v dialógu. Tieto koncové body sú definované ako miesta relevantné pre zmenu rečníka (Transition-Relevance Place, TRP), čo značí možné miesto prechodu od jedného rečníka k druhému. Avšak treba podotknúť, nie na každom konci TCU musí dôjsť k prechodu. Z tohto dôvodu sa tieto body charakterizujú ako relevantné, ale nie nutné pre prechod. Výskyt a využitie bodov TRP v dialógu zabezpečuje zachovávanie plynulého striedania sa rečníkov. TRP teda uľahčuje každému účastníkovi rozpoznať, kedy bude môcť začať alebo ukončiť konkrétne TCU. Z tohto dôvodu musí byť TRP jasne predvídateľné pre každého poslucháča, aby sa dosiahol hladký prechod medzi rečníkmi. Výsledkom takéhoto prechodu je minimalizovanie medzier a prekryvia medzi jednotlivými prehovormi. Komplikujúcim faktorom v skúmaní miest prechodov medzi rečníkmi je, že TRP nemožno priamo pozorovať v dátach. Môžeme pozorovať iba skutočné obraty, ktoré by sa teda

mohli považovať za podmnožinu TRP. Avšak obraty sa môžu vyskytovať aj tam, kde neexistuje TRP [3].

Dôležitým pojmom pri analýze striedania iniciatívy je spätná väzba (Back-Channel, BC). Spätnú väzbu možno charakterizovať ako výrok alebo prehovor účastníka v roli poslucháča, avšak bez úmyslu prebrať iniciatívu v rozprávaní od aktuálneho rečníka (bez úmyslu prevziať rolu rečníka). Tieto signály spätnej väzby slúžia väčšinou na prejavenie súhlasu, porozumenia a tiež povzbudzujú aktuálneho rečníka, aby pokračoval (continuers). Spätná väzba sa zvyčajne vyskytuje počas krátkych páuz v prehovore rečníka a zvyčajne trvá menej ako sekundu. Výroky BC primárne signalizujú neustálu pozornosť, súhlas alebo rôzne emocionálne reakcie. Spätná väzba má špeciálny status v rámci pravidiel striedania iniciatívy, pretože je pomerne často sa vyskytujúca počas konverzácie, no napriek tomu sa nepovažuje za miesto vhodné pre zmenu rečníka. Podobne ako sa klasické striedanie rečníkov vyskytuje po určitých náznakoch v reči, aj načasovanie spätnej väzby súvisí s určitými podnetmi, ktoré ju naznačujú. Tieto podnety nazývame „backchannel inviting cues“. Miestá, ktoré nasledujú tesne po týchto podnetoch vnímame ako relevantné pre umiestnenie spätnej väzby aktuálnym poslucháčom a označujeme ich ako miesto relevantné pre spätnú väzbu (Backchannel-relevant place, BRP) [1],[4].

Aby bolo možné čo najjednoduchšie identifikovať TCU a TRP, výskumníci zaoberajúci sa rečovými technológiami zistili, že je vhodné segmentovať reč do medzi-pauzových jednotiek (Inter-Pausal Unit, IPU). Ide o úseky hovorenej reči od jedného rečníka bez akéhokoľvek ticha presahujúceho určitú hodnotu (napr. 200 ms). Prehovor alebo obrat je potom typicky definovaný ako sekvencia IPU od jedného rečníka, ktoré nie sú prerušované jednotkami IPU od iného rečníka. Ticho medzi dvoma IPU od toho istého rečníka sa zvykne označovať ako pauza (pause), zatiaľ čo ticho medzi IPU od rôznych rečníkov sa označuje ako medzera (gap) [1].

II. PROZÓDIA REČI V RÁMCI STRIEDANIA INICIATÍVY

Úloha prozódie v rámci striedania iniciatívy sa stala predmetom veľkého záujmu aj sporov. Prozódia sa vzťahuje na neverbálne aspekty reči vrátane intonácie, hlasitosti, rýchlosti reči a podobne. Zistilo sa, že slúži mnohým dôležitým funkciám v konverzácii, vrátane vyhodnocovania dôležitosti, syntaktickej jednoznačnosti, postojových reakcií, neistoty alebo presunov v rámci tém. Pokiaľ ide o intonáciu, štúdie naprieč rôznymi jazykmi zistili, že stabilná úroveň intonácie (v strede základného frekvenčného rozsahu hovoriaceho) blízko konca IPU má tendenciu slúžiť ako náznak pre zadržanie obratu (turn-holding), čiže dialógovej výmeny medzi rečníkmi. Zatiaľ čo stúpajúca alebo klesajúca výška tónu sa dá nájsť v kontextoch, v ktorých dochádza k náznakom uskutočnenia obratu (turn-yielding) medzi rečníkmi.

Intenzita hlasu nesie taktiež istú informačnú hodnotu pre detekciu miest TRP. Rečníci majú tendenciu stlmiť hlas, keď sa blížia k hraniciam potenciálneho obratu, zatiaľ čo pri pauzách vznikajúcich vo vnútri prehovoru mala reč vyššiu intenzitu. Niekoľko štúdií taktiež skúmalo akú rolu hrá prozódia pre vyvolanie spätnej väzby a ako sa tieto náznaky líšia od náznakov pre uskutočnenie obratu. Zistili, že spätná väzba má tendenciu prichádzať asi 200 ms po oblasti s nízkym tónom. Na druhej strane taktiež zistili, že IPU bezprostredne pred spätnou väzbou vykazovali jasnú tendenciu ku konečnej stúpajúcej intonácii, ako aj vyššej intenzite. Tieto trochu protichodné zistenia možno vysvetliť jazykovými rozdielmi v rámci vykonaných štúdií. Je však nutné poznamenať, že v žiadnej dostupnej štúdií nebola realizovaná analýza dát v slovenčine.

Bez ohľadu na úlohu prozódie pri striedaní iniciatívy medzi ľuďmi môže prozódia poskytnúť dôležité podnety z pohľadu konverzačného systému. Keďže konverzačné systémy nemajú rovnaké výpočtové/kognitívne obmedzenia a nemusia si vopred pripravovať odpoveď v takom rozsahu ako ľudia, mohli by vo väčšej miere využívať koncové signály [1].

A. Výška tónu a základná frekvencia

Výšku tónu (pitch) možno charakterizovať ako percepčný atribút zvuku. Fyzikálnym korelátom výšky tónu je základná frekvencia (fundamental frequency, F0) vibrácií hlasiviek. Základná frekvencia je definovaná ako priemerný počet oscilácií za sekundu a býva vyjadrená v hertzoch [Hz]. Táto frekvencia je špecifická pre každého rečníka kvôli rozdielom vo fyzickej štruktúre hlasiviek rečníka. Typické hodnoty základnej frekvencie ležia zhruba v rozsahu od 80 až po 450 Hz. Dynamika obálky frekvencie F0 je ovplyvnená niekoľkými faktormi, ako je identita hovorenej zvukovej jednotky, poloha vzhľadom na frázu alebo slová, kontext, štýl reči konkrétneho rečníka, intonačné pravidlá jazyka, typ vety (opytovacia alebo oznamovacia), atď. Základná frekvencia je jedným z hlavných parametrov, pomocou ktorého je možné rozlíšiť pohlavie hovoriaceho. F0 je taktiež dôležitým znakom, ktorý charakterizuje jednotlivých rečníkov a ich emocionálny stav [5].

B. Intenzita

Intenzita (intensity) reči je všeobecne uznávaná ako jeden z troch základných prozodických parametrov. Pojem intenzita býva často nahrádzaná pojmi amplitúda alebo hlasitosť. Intenzita je základným prvkom amplitúdy a je definovaná ako sila prenášaná zvukovými vlnami na jednotku plochy. Sluchové vnímanie intenzity býva spravidla vyjadrené v decibeloch [dB]. Väčšina lingvistov však nevenuje intenzite ako charakteristike významnú pozornosť. Intenzita hrá významnú úlohu pri definícii slabiky, ktorá hovorí, že slabika zodpovedá vrcholu intenzity. Intenzita je demarkačná funkcia na rôznych úrovniach. Hodnota intenzity počas daného časového intervalu môže byť použitá pre detekciu páuz a tak môže oddeľovať reč od nerečových sekvencií. Taktiež vzájomný vzťah medzi intenzitou a základnou frekvenciou možno považovať za určitý fyziologický základ. Intenzita a základná frekvencia sú riadené rovnakými mechanizmami, ako je zvýšenie pľúcneho úsilia a subglotálneho tlaku, napätie hlasiviek, atď. Dá sa predpokladať, že vyššia základná frekvencia je vo všeobecnosti korelovaná so zvýšením intenzity. Podobne je pokles F0 na konci viet spojený s poklesom intenzity [6].

III. AUTOMATICKÁ DETEKCIA MIEST TRP

A. Dialógový korpus

Základom dialógového korpusu, z ktorej vychádzame v rámci výskumu v tejto práci, sú audio záznamy rozhovorov resp. diskusií prebiehajúcich medzi dvoma rečníkmi. Konkrétne pracujeme so zvukovými nahrávkami diskusnej relácie Pod lampou, kde vystupujú dvaja rečníci (moderátor a hosť). Pri spracovávaní dát vychádzame zo záznamov vo forme wav súborov. V tejto práci sa zameriavame na analýzu prozodických vlastností reči jednotlivých účastníkov v rámci striedania iniciatívy medzi rečníkmi a taktiež ich možné využitie pri detekcii miest TRP. Potrebné prozodické charakteristiky postupne extrahujeme z časti reči konkrétnych účastníkov. Za týmto účelom dostupné nahrávky spracovávame do vhodnej podoby najprv v programe Audacity a následne v softvérovom balíku Praat. V poslednom kroku spracovania databázy dát použijeme softvér Microsoft Excel pre zlúčenie a usporiadanie všetkých získaných dát.

V softvéri Audacity pri postupnom počúvaní rozhovoru vystrihujeme konkrétne úseky nahrávky s dĺžkou trvania cca 1,52 sekundy. Tieto jednotlivé úryvky majú dvojakú formu. Prvým typom sú úryvky reči, po ktorej dochádza k vzniku miesta TRP a druhým typom sú úryvky reči, po ktorej nenastáva vznik miesta TRP. Dĺžka úryvku pozostáva z 1,32 sekundového okna reči tesne pred TRP nasledovaným krátkym úsekom ticha (medzery) o dĺžke 0,2 sekundy t. j. 200 ms. Hodnota 1,32 sekundy vychádza zo zistenia prof. Beňuša (2016) [7], ktorý tvrdí, že priemerné trvanie IPU v slovenskom jazyku predstavuje práve túto hodnotu. Štandardnú dĺžku ticha resp. čas prechodu medzi rečníkmi definuje množstvo štúdií približne ako hodnotu 200 ms. Práve v rámci tejto medzery väčšinou dochádza k jasnému vzniku miest TRP, avšak táto medzera sa nemusí vyskytovať pri všetkých prípadoch vzniku TRP. Niekedy je striedanie medzi rečníkmi rýchle bez badateľnej medzery medzi nimi. Oba typy úryvkov extrahujeme z reči každého účastníka v rámci jedného približne hodinového rozhovoru. Polovica získaných úryvkov predstavuje miesta „ano_TRP“, čiže úseky reči, po ktorých vzniká miesto TRP. Druhá polovica úryvkov predstavuje miesta „nie_TRP“, čiže úseky reči, po ktorých nevzniká miesto TRP. Takýmto postupom ručne pripravíme databázu úryvkov pre každého účastníka konverzácie, s ktorými budeme ďalej pracovať.

Program Praat využívame pre získanie potrebných charakteristík reči z každého úryvku získaného v predošlom kroku. V našom výskume sa konkrétne zameriame na extrakciu dvoch prozodických charakteristík reči, a to intenzity (intensity) a výšky tónu (pitch) rečového signálu. Postup extrakcie charakteristík pozostáva z niekoľkých krokov, ktoré realizujeme pre každý získaný úryvok reči postupne od všetkých rečníkov.

Ďalším potrebným krokom pre náš nasledujúci výskum je zlúčenie a usporiadanie získaných dát do vhodnej podoby pre nasledujúcu prácu s nimi. Pre túto úlohu sme zvolili tabuľkový softvér Microsoft Excel z balíku Microsoft Office. Postupne sme nakopírovali získane hodnoty intenzity a výšky tónu každého úryvku reči, ku ktorým sme priradili označenie v podobe znaku „A“ pre úryvky obsahujúce miesta TRP alebo znak „N“ pre úryvky, ktoré neobsahujú miesta TRP. Jednotlivé hodnoty sme umiestnili v postupnosti pod seba do riadkov tabuľky. Najprv všetky úryvky obsahujúce miesta TRP, pod to úryvky neobsahujúce TRP. V stĺpcoch sa najprv nachádzajú hodnoty pre intenzitu, ďalej nasledujú stĺpce s hodnotami výšky tónu. Na záver sa na konci nachádza stĺpec obsahujúci prislúchajúcu hodnotu TRP (A alebo N). Takto zlúčené dáta sme na záver rozdelili do samostatných tabuliek, kde 75% zo všetkých dát použijeme na tréning klasifikátorov a zvyšných 25% dát neskôr použijeme pri testovaní jednotlivých natrénovaných klasifikačných modelov.

B. Automatická detekcia miest TRP

V rámci nášho výskumu sme sa rozhodli zamerať na využitie metódy klasifikácie, ktorá spadá pod strojové učenie pod dohľadom. Táto metóda využíva označené vstupné dáta pre realizáciu tréningu a testovania klasifikačných modelov. Cieľom klasifikácie je naučiť sa klasifikačné pravidlo, na základe ktorého by bolo možné s určitou presnosťou ďalej realizovať automatické určovanie dát. Pre účely nášho výskumu bolo najprv potrebné pripraviť vhodnú databázu dát, s ktorými budeme pracovať. Finálne spracované dáta pozostávajú z extrahovaných prozodických parametrov reči niekoľkých rečníkov. V ďalšej časti práce tieto dáta použijeme pre automatické tréningovanie a následné testovanie rôznych klasifikačných metód. Jednotlivé klasifikačné metódy aplikujeme na dáta prostredníctvom automatizovaných klasifikátorov dostupných v nástroji Classification Learner softvérového balíka Matlab. Ako vstupné dáta sme použili údaje získané z reči každého rečníka, najprv jednotlivo a následne kombináciu dát od viacerých rečníkov súčasne.

1) Classification Learner

Nástroj Classification Learner patrí k dostupným aplikáciám balíka Matlab spoločnosti MathWorks. Tento nástroj ponúka automatizované tréningovanie modelov na klasifikáciu údajov. Pomocou tejto aplikácie je možné skúmať strojové učenie pod dohľadom pomocou rôznych klasifikátorov. Umožňuje skúmať zvolené údaje, vyberať funkcie, špecifikovať schémy validácie, modely tréningovania a vyhodnocovať získané výsledky. Classification Learner vykonáva automatizované tréningovanie pre vyhľadanie najlepšieho typu klasifikačného modelu vrátane rozhodovacích stromov, diskriminačnej analýzy, podporných vektorových strojov, logistickej regresie, klasifikátorov najbližšieho suseda, naivných bayesových klasifikátorov, atď. Strojové učenie pod dohľadom umožňuje vykonávať automatickú klasifikáciu zadaním známeho súboru vstupných údajov (pozorovaní alebo príkladov) a známých odpovedí na údaje (napr. štítky alebo triedy). Zadané vstupné údaje sa použijú na tréningovanie modelu, ktorý generuje predikcie pre odozvu na nových údajoch. Ďalej umožňuje zvolený natrénovaný model exportovať do pracovného priestoru Matlab alebo vygenerovať Matlab zdrojový kód na opätovné vytvorenie natrénovaného modelu [8].

2) Tréningovanie a testovanie modelov

Pre účely automatickej klasifikácie miest TRP v hovorenom dialógu sme realizovali tréningovanie a testovanie klasifikátorov s využitím dát od troch rečníkov, ktoré pozostávali z extrahovaných prozodických charakteristík reči každého rečníka. Automatické tréningovanie klasifikátorov sme uskutočnili najprv pre každého rečníka samostatne. Neskôr sme skúmali výkonnosť automatickej klasifikácie aj pre prípad, kde vstupné dáta predstavujú kombináciu dát od týchto rečníkov. Najprv sme využili kombináciu vstupných dát od prvých dvoch mužských rečníkov, potom sme k dátam pridali tretieho ženského rečníka. Pozorovali sme ako sa budú presnosti klasifikátorov meniť vzhľadom na zvyšujúce sa množstvo a rôznorodosť vstupných dát.

Pri tréningovaní všetkých dostupných klasifikátorov, nástroj Classification Learner priebežne vyhodnocuje validačnú presnosť (Accuracy), na základe ktorej je možné získať okamžitú predstavu o presnosti klasifikácie každého modelu. Pre každý druh použitého klasifikátora nástroj vykonáva tréningovanie viacerých typov daného klasifikátora. Pre účely nášho výskumu pre každý druh klasifikátora vyberieme typ, ktorý dosiahol najlepšiu validačnú presnosť. Validácia presnosť sa vypočítava na súbore údajov, ktorý sa nepoužíva priamo pri tréningovaní, ale využíva sa (počas tréningového procesu) na predbežné overenie výkonnosti modelu.

Pre každý natrénovaný model sme v prostredí Matlab realizovali hodnotenie presnosti modelu taktiež charakteristikami „precision“, „recall“ a „F-score“. Tieto funkcie slúžia na vyhodnocovanie predikčnej presnosti každého modelu. V oblasti rozpoznávania vzorov, získavania informácií a klasifikácii (strojové učenie) predstavujú *precision* a *recall* metriky výkonnosti, ktoré sa vzťahujú na údaje získané zo súboru dátových vzoriek. F-skóre meria presnosť klasifikácie, pričom kombinuje obe metriky *precision* a *recall*.

IV. DOSIAHNUTÉ VÝSLEDKY

V Tab. 1 sú súhrne zaznamenané výsledky presnosti jednotlivých typov klasifikátorov pre všetky prípady tréningovania. Pre prípad 1. rečníka dosiahli totožný najlepší výsledok 98,4% klasifikátor rozhodovacích stromov (Decision Tree) a súborový (Ensemble) klasifikátor. Pre prípad klasifikácie 2. rečníka dosiahol najlepší výsledok 92,7% klasifikátor Ensemble. Takisto pre prípad 3. rečníka dosiahol najvyššiu presnosť 94,0% klasifikátor Ensemble. V prípade klasifikácie 1. a 2. rečníka súčasne dosiahol najlepší výsledok 93,6% klasifikátor Ensemble. Na záver pre prípad klasifikácie 1., 2., a 3. rečníka súčasne dosiahol najlepší výsledok 94,4% taktiež klasifikátor Ensemble. Suverénne najlepšie výsledky dosahoval klasifikátor Ensemble, ktorý využíval súborovú metódu náhodný les (Random forest). Táto metóda strojového učenia pre

klasifikáciu funguje tak, že v čase tréovania vytvára množstvo rozhodovacích stromov. Pre klasifikačné úlohy je teda výstupom náhodného lesa trieda (class), ktorú vyberie väčšina rozhodovacích stromov.

Tab. 1
Výsledky automatickej klasifikácie TRP (presnosť – ACC)

Typ klasifikátora	1. rečník	2. rečník	3. rečník	1. a 2. rečník	1.,2.,3. rečník
Decision Tree	98,4%	86,2%	92,9%	93,4%	92,3%
Linear Discriminant	64,1%	84,9%	87,8%	88,2%	92,3%
Logistic Regression	64,1%	81,2%	85,5%	83,8%	85,8%
Naive Bayes	95,3%	90,4%	91,8%	91,0%	90,8%
SVM	96,1%	91,7%	92,9%	92,5%	92,8%
KNN	92,2%	90,8%	91,2%	89,9%	92,0%
Ensemble	98,4%	92,7%	94,0%	93,6%	94,4%

V. ZÁVER

V predloženej práci sme sa zamerali na automatickú detekciu miest relevantných pre zmenu rečníka, tzv. TRP pomocou algoritmov strojového učenia. Na natréovanie klasifikátorov sme použili rečový dialógový korpus, ktorý pozostáva z dialógov v slovenčine a k nim vytvoreným prepisom s označením TRP bodov. Vstupné dáta sme parametrizovali dvojicou prozodických parametrov – vektormi hodnôt F0 a intenzity, ktoré boli vypočítané po rámcoch s dĺžkou 1,52ms. Natréované klasifikátory s limitovaným objemom dát pre tréovanie aj testovanie dokázali detegovať rámce s výskytom TRP bodu s presnosťou viac ako 90%, čo poukazuje na vhodnosť prozodických parametrov na túto úlohu. Môžeme konštatovať, že prozódia hrá dôležitú úlohu v signáloch súvisiacich so striedaním rečníkov v dyadických interakciách.

POĎAKOVANIE

Táto práca bola podporená Agentúrou na podporu výskumu a vývoja vďaka projektu APVV SK-TW-21-0002, Vedeckou grantovou agentúrou realizáciou výskumného projektu VEGA 1/0753/20 „Robustné rečové technológie metódami hlbokého učenia“ a VEGA 2/0165/21 „Technológie automatického spracovania reči na pomoc v krízových situáciách“ financovaných z prostriedkov Ministerstva školstva, vedy, výskumu a športu SR.

LITERATÚRA

- [1] G.Skantzé, “Turn-taking in Conversational Systems and Human-Robot Interaction: A Review”, Computer speech & language (Print), vol. 67. Elsevier BV, 2021.
- [2] J. Holler, K. H. Kendrick, M. Casillas, S. C. Levinson, “Editorial: Turn-Taking in Human Communicative Interaction”, *Frontiers in psychology*, 6, 1919. <https://doi.org/10.3389/fpsyg.2015.01919>, 2015
- [3] S. E. Clayman, “Turn-constructive units and the transition-relevance place”, In *The Handbook of Conversation Analysis* (Jack Sidnell, Tanya Stivers, eds.), Oxford, U.K., Wiley-Blackwell, pp. 150–166, 2012
- [4] M. Kuswandi, Y. Apsari, “An analysis of pauses, overlaps and backchannels in conversation in vlog by Nessie Judge”, *Project (Professional Journal of English Education)*. 2. 282. 10.22460/project.v2i3.p282-291, 2019
- [5] M. Sigmund, “Statistical Analysis of Fundamental Frequency Based Features in Speech under Stress”, *Information Technology And Control*. 42. 10.5755/j01.itc.42.3.3895, 2013
- [6] E. Koffi, “A comprehensive review of intensity and its linguistic applications”, *Linguistic Portfolios*, 9(1), 2., 2020
- [7] A. Gravano, P. Brusco, S. Benus, “Who Do You Think Will Speak Next? Perception of Turn-Taking Cues in Slovak and Argentine Spanish”, 1265-1269. 10.21437/Interspeech.2016-585, 2016
- [8] Classification Learner. [Online] [cit. 2022-02-08] Dostupné na internete: <<https://www.mathworks.com/help/stats/classificationlearner-app.html>>

Portal for automated quality assessment of goods

¹*Patrik SCHWARTZ*, ²*Ondrej KAINZ*, ³*Miroslav MICHALKO*,
⁴*František JAKAB*, ⁵*Ivana NOVÁVKOVÁ*

^{1,2,3,4,5}Department of Computers and Informatics, Faculty of Electrical Engineering and Informatics Technical University of Košice, Slovak Republic

¹patrik.schwartz@student.tuke.sk, ²ondrej.kainz@tuke.sk, ³miroslav.michalko@tuke.sk,
⁴frantisek.jakab@tuke.sk, ⁵ivana.novakova@tuke.sk

Abstract – The aim of this paper is to create a web portal that would provide the user with the ability to recognize the type of goods and their quality using the detection of objects in the image. Based on a comparative analysis of existing object recognition algorithms, the YOLO v3 algorithm was selected. Subsequently, standardized procedures for evaluating the quality of selected goods, collector coins, were analyzed. Thanks to this knowledge, the critical points and features that needed to be addressed in the design were identified and the quality levels that the coin must meet were determined. The analysis of a suitable visual input to the system, from the point of view of exposure, determined the requirements for the minimum quality of coin photos for their further processing. Last but not least, image processing techniques to obtain the necessary properties were explained. The synthetic part of this work describes the design of three parts of this solution: neural network training, web portal consisting of client and server part, and API, which will work with the exported neural network model and provide functionality of object detection and quality assessment. The results of this work were finally experimentally evaluated in terms of the usability of the portal and the efficiency of the implemented algorithms using standardized measurements. The user test group identified critical elements of the portal that needed to be modified. Based on the comparison of object detection results on datasets of different quality, the quality of the trained neural network was determined and finally the accuracy of our solution was determined by comparing the results of quality evaluation with real results from other web portals.

Keywords – neural network, object detection, recognition, classification, quality assessment, web portal

I. INTRODUCTION

In today's technological world, it is already common practice to establish, in addition to the brick-and-mortar store, an online store, which customers will always have at hand and will be able to search the product catalog and shop from the comfort of their home. A special category of such online stores are portals on which customers are able not only to monitor the current offer of products, choose from them and buy, but also to contribute with their own products or products and sell them to other users. In some cases, these are portals where the user adds items that are no longer in 100% condition, but still have a high price. These are mainly collectibles, which can have great value even in a worse condition, which increases with the quality of the product and its rarity. To determine such wear, various types of collector's items use various methods to help the owner assess the quality of the individual items in his collection. The prerequisite for assessing the condition of a collector's item is first the recognition of its type. all types of items. This work is therefore focused on a specific group of collectibles, namely collector coins, which are among the most frequently collected items around the world.

II. ANALYSIS OF OBJECT RECOGNITION AND EVALUATION IN THE IMAGE

Automated quality assessment requires several necessary steps. The first of them is the detection of the object in the picture. This chapter will therefore focus on the comparative analysis of algorithms for the individual steps of determining the quality of goods described above.

A. Comparative analysis of algorithms for object detection in the image

The first step to evaluating an object is its detection. Today, image detection is used in countless industries, but also in everyday life. Object detection consists of a combination of object location and object classification. It is the localization of the object in the image that increases the complexity of this issue, because on the basis of these results, the second mentioned step takes place, namely the classification of the object. The first step to evaluating an object is its detection. Today, image detection is used in countless industries, but also in everyday life. Object detection consists of a combination of object location and object classification. It is the localization of the object in the image that increases the complexity of this issue, because on the basis of these results, the second mentioned step takes place, namely the classification of the object. Several algorithms are based on this principle. The following section will describe the most used algorithms for finding the object in the image, from which the most optimal algorithm for our case of collector coins will be selected on the basis of comparative analysis.

1) *Fast R-CNN*: Fast R-CNN is an abbreviated name for the fast regional convolution network method for object detection. As the author states in his work [1], it is based on the R-CNN algorithm and its task is mainly to improve the speed of training and testing and at the same time increase the accuracy of detection. R-CNN achieves excellent object detection accuracy using a deep convolution network system for object design classification.

2) *Faster R-CNN*: Faster R-CNN is a modified version of Fast R-CNN. As stated by the author [2], the main difference between them is precisely in the mentioned generation of areas of interest. As mentioned above, Fast R-CNN uses selective search to generate areas of interest, while faster R-CNN uses so-called RPN, i.e. the Region Proposal Network. RPN takes as a pixel map input and generates a set of object designs, each of which outputs an objectivity score. Figure ?? shows the architecture of the Faster R-CNN algorithm. As with Fast R-CNN, the input here is an image, which is passed to a fully convolution network, which returns a map of features for that image. The aforementioned RPN is applied to these property maps, which returns object designs along with their objectivity scores. A link layer of areas of interest is applied to these designs to reduce designs to the same size. Finally, these modified designs are passed to a fully interconnected network, which has a softmax layer at its apex and a linear regression layer for classifying and returning bounding boxes for individual objects.

3) *SSD*: Real-time object detection in the image is not very possible with Fast R-CNN and Faster RCNN algorithms, as their processing takes longer than necessary. In his publication [3], the author proposes the so-called SSD (Single Shot Multibox Detector) algorithm that solves this problem and can be used to detect objects in real time. The SSD algorithm achieves higher processing speeds by eliminating the need for regions of interest. Instead, it uses other processes to match the accuracy of the previous two algorithms. Accuracy is increased with multi-scale functions and preset boxes.

4) *YOLO*: Another algorithm that is used to detect objects in the image is the so-called YOLO (You Look Only Once) algorithm, which does not use areas of interest to locate objects in the image, but uses the entire image in one instance and predicts the coordinates of the bounding box and class probabilities for these fields. It's incredibly fast and can process 45 frames per second. In the next section, we will describe the steps of this algorithm. Like previous algorithms, YOLO takes an unprocessed and unmodified image as input, which is then divided into a grid, e.g. with dimensions of 3 x 3. Image classification and localization are applied for each cell of this grid. YOLO then predicts bounding boxes and their corresponding class probabilities for objects, if such objects were found. As the author states in his publication [4], unlike techniques based on sliding window and area design, YOLO sees the whole picture during training and testing, so it implicitly encodes contextual information about the classes as well as their appearance. A method like R-CNN makes mistakes because it doesn't see more context. YOLO generates less than half the number of background errors compared to Fast R-CNN. YOLO learns generalizable representations of subjects. When trained on natural images and tested on works of art, YOLO far outperforms cutting-edge detection methods such as R-CNN. Because YOLO is a highly generalizable algorithm, it is less likely to fail when applied to new domains or unexpected inputs.

5) *Algorithm comparison*: Since each algorithm has its advantages and disadvantages, it is necessary to determine which algorithm best meets the requirements of the proposed system. From the customer's point of view, the highest possible speed of image prediction and evaluation is required, while efficiency is just as important. In Figure 1, a comparison of detection algorithms based on the frequently used COCO mAP 50 benchmarking is shown. As can be seen in Figure 1, YOLO v3 works at the same level as other state-of-the-art detectors, while being significantly faster. Compared to the SSD algorithm, it shows slightly better results, but is significantly faster. Based on these findings, the YOLO v3 algorithm was therefore selected in the case of this work.

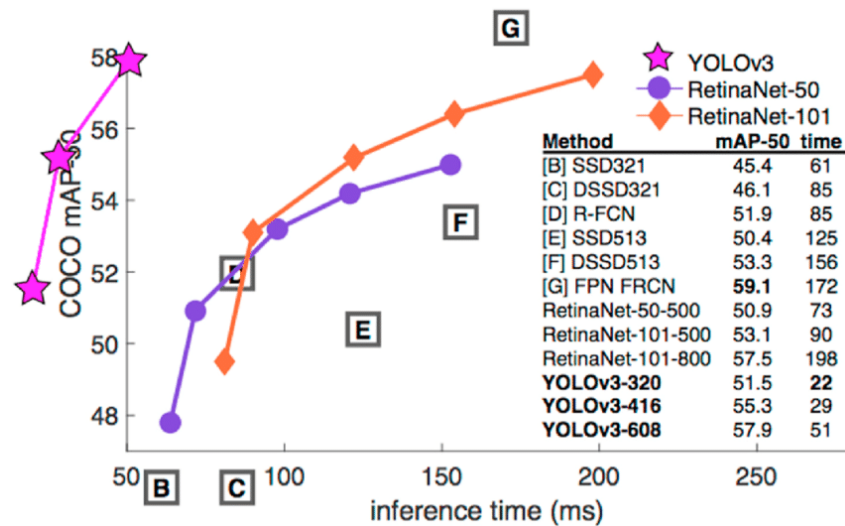


Fig. 1 Comparison of detection algorithms using COCO mAP 50 benchmarking [5]

B. Analysis of properties extraction properties

As stated on the NGum Numismatic Guaranty Company website [6], the company uses a numerical level to briefly describe the smooth condition of the coins. Available numerical levels range from 1 to 70 based on an internationally recognized scale. A higher grade indicates the better quality of the coin, the lowest grade the worst quality. Thus, the evaluation of a coin passes through this scale, stopping at a level where the coin does not meet the relevant requirement.

Before the training of the neural network, it is necessary to effectively extract the properties according to which the detection and recognition of the object in the image will take place. In the case of collector coins, these assumptions must be taken into account:

- coins are usually round in shape, but in the case of older roundings may be broken by damaging the edges of the coin,
- the coin can be rotated in any way, but the pattern must be recognized at any angle,
- photographer and can be made under a certain bevel, which can degrade the shape and structure of the coin,
- the coin is placed on a horizontal base, which can be of different colors and textures,
- the coin can cast a shadow in different directions and sizes, so it needs to be removed in advance.

As the author states in his book [7], coin collectors must adhere to certain rules or advice when creating coin photographs in order to preserve and clearly identify contours, cutouts as well as dirt and grain on the coin. The author states that only bad lighting can lose most of the coin's data, which would otherwise be crucial for recognition. The author deals with the recognition of coins using computer technology in his publication [8]. In the following chapters, therefore, the most common algorithms for editing the input image and for extracting the necessary data will be discussed.

1) *Image normalization*: As stated by the author in his publication [9], data normalization is a key step towards obtaining a stable dataset and good results in training and later use of the neural network in practice. Without image normalization, the dataset would consist of photographers that would be too different, either in quality or exposure. The process of normalization of the coin image, which the author proposes in his work [10].

2) *Convert a color image to grayscale*: According to the author [11], the method of transforming a color RGB image into grayscale is a common and recommended method for detecting and recognizing objects in an image. The author emphasizes [12] that the transformation of the image to grayscale reduces the computational complexity by reducing the dimensionality typical of the transformation of the RGB values of the tristimulus (R, G, B) to the intensity values (I). At the same time, this emphasizes the contours of the coin, which will be needed to determine its type and quality.

3) *Shadow removal*: The shadow left in the image after adjusting to shades of gray could be the reason for the inaccuracy and misclassification.

The author in his work [13] assumes that coins are always circular in shape. Based on this hypothesis, the Hough Transform for Circle Detection (CHT) is used to remove the shadow, which is used to find circular patterns in the image. It is used to transform a set of point elements in an

image into a set of accumulated voices in the parameter space, which are stored in the form of a field, and the highest number of fields indicates the presence of a shape. The CHT algorithm requires 3 parameters, namely the x, y coordinates of the center of the circle and the radius.

DESIGN OF A WEB SOLUTION

The proposed web solution should be in the form of a portal that will provide electronic commerce services (e-commerce). For the needs of the final solution of this work, it is necessary to ensure the following system requirements:

- product management for portal users - adding, editing, deleting own products,
- a message for the site administrator that offers more privileges than for other users,
- arranging electronic commerce for the sale and purchase of collector coins,
- sufficient memory storage for storing product images and temporal quality files,
- clear and user-friendly interface,
- guarantee the compatibility of all tools and technology used.

The proposed web portal will consist of catalogs and products added by users. In order to be able to control this process with regard to quality and security, it is necessary to ensure that only registered and logged-in users whose identities have been verified on the basis of an activation email during registration are able to add content to the portal. The registered and authenticated user is then allowed to manage their product catalog. He is allowed to add, edit or delete products in his catalog and is able to sell them to other users, respectively. buy new. When adding a new product, the user must go through a quality evaluation process in order to be able to categorize this product and set the detected quality level. In Figure 2 you can see an activity diagram describing the main processes when adding a new product.

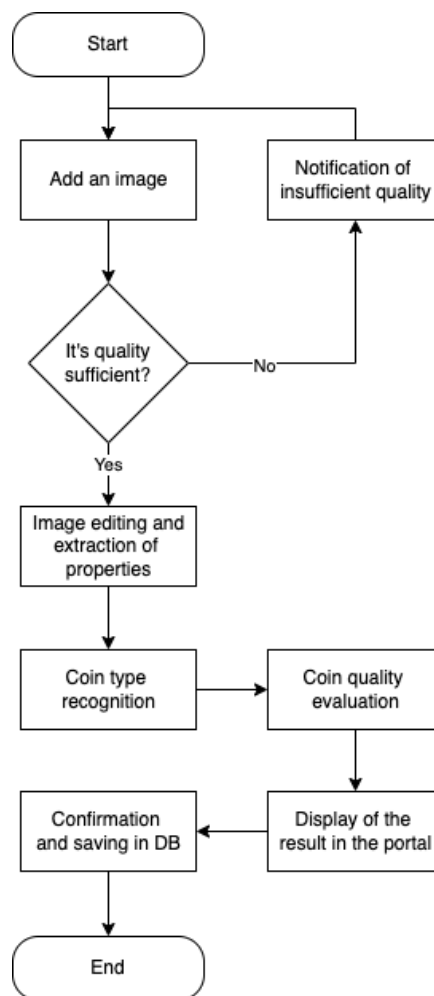


Fig. 2 Abstract activity diagram of adding goods to the portal

The quality assessment process will return the result based on implemented numismatic standards described above. A closer description of this process will be described below. Subsequently, the result will be displayed as a user response along with other coin information obtained. Using

a detecting algorithm and correctly determined classes in classmakers, it will be possible to assign the product to the corresponding color, value, origin and other hairyess that each product will have.

TRAINING OF A NEURAL NETWORK

It is possible to see the sequential diagram of the basic steps of creating YOLO v3 neural network. The first step is to obtain the largest and most diverse dataset of collector coins. The larger the set is available, the greater the possibility that the neural network will be properly trained. If the dataset was too small, there would be a so-called. Under -fitting. This may also occur if we set a very low number of epoch when training a neural network, causing the neuron network to be able to learn all the information that our solution requires. The opposite case may be so-called. Overfitting. In such a situation, the neuron network can learn too much information that can be unnecessary, disturbing or even undesirable, which can lead to poor classification of some cases. The distribution of the dataset also plays an important role in the training and test set. However, as the author states in his book [14], in recent years, the dataset layout for training, validation and testing undercoats has also been used.

This adjustment is followed by training the neural network and optimization of hyperparameters using a validation set to achieve the best results. The training itself is preceded by the correct initialization of the neural network, the choice of optimization, the gradient and the calculation of the scales after each step. Only after the completion of the convention can be completed to the training itself. After training of the neural network, a validation is carried out, after which, if necessary, the parameters are modified. Subsequently, the trained and rolled model on the test set is tested. If the model is returned to the dilapidate of prediction results, a model is exported, which will then be recorded to the server storage.

QUALITY EVALUATION

A technique that is also used in numismatic circuits is used to evaluate the quality of a coin, namely a comparison with a coin that has minimal damage and therefore the highest quality. With this, it is possible to find out which edges are unwanted, which are destroyed, which are abraded. We can also determine the original and correct color intensity, gloss of the material and other properties. However, since it is difficult to find a coin that is in a sufficiently preserved condition for many coins, it is necessary to simulate such a condition. This can be achieved by various techniques that remove most of the coin's deficiencies and create an artificially created, more refined version of the coin, with which the original coin is finally compared. This procedure has one advantage, and that is in making a copy. Comparing with such a copy will be easier than with a completely different image of a coin that should be cropped, rotated, enlarged to fit every single pixel of a coin with a controlled coin.

IMPLEMENTATION

The aim of this work is to create a web solution that will provide the implemented functionality of goods quality evaluation, specifically on the production web portal, which will provide the basic elements of e-commerce. This solution will therefore consist of several architectural parts. In Figure 3 it is possible to see an abstract representation of the technologies used in individual parts of the system architecture and individual actions, resp. relations between them.

The solution itself therefore consists of a client part with which the user interacts. Based on GET and POST HTTP requests, data is obtained from the server part. It is responsible not only for creating pages with the right content, but also for communication with the database, memory storage, payment gateway and also with the implemented system for prediction and evaluation of goods.

After logging in, the user is able to create their own coin catalogs, and needs to enter the name and description of this catalog. Following deposit, he is allowed to add coins to this catalog. Adding a coin takes place in several steps. First, the user is asked to upload a coin image. After checking the image quality, the image is stored in the cloud storage and an instance of the product is created, which is stored in the database, and it is set to be non-active on the website. This means that such a product, which is in the process of being created, is not visible to other users of the portal. Subsequently, requests for implemented APIs are sent sequentially. They return the predicted coin type and the rated coin quality. Based on the output of these requests, the user is shown a product editing page to which data has already been assigned based on the predicted coin type and is also assigned the quality category to which it belongs. At the same time, he will see a picture of the coin, which highlights the shortcomings and damage found.

Using the FastAPI development framework, POST endpoints are easily created for:

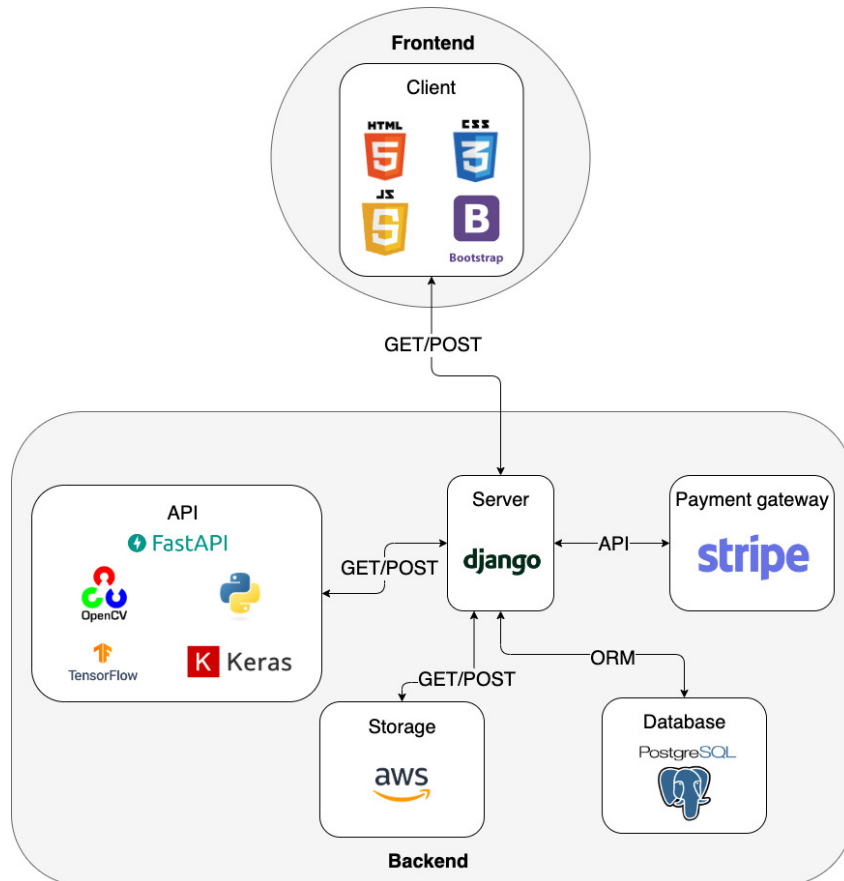


Fig. 3 System architecture

- coin type recognition,
- quality evaluation.

The following section will describe the implementation of each API endpoint.

Coin type recognition

To start the classification, the endpoint of the implemented API / net / image / prediction / is called, which has the only necessary argument, namely the path to the image. After checking the input image, the image is loaded into the working memory and transferred to the field numpy, which allows us to determine the size of the image, but also other information such as. color scale. Field conversion is also advantageous in terms of computational efficiency. Using the implemented functions, the desired image adjustment is performed. The built-in predict () function is then called to return the prediction boxes. As explained in the analytical part, the YOLO v3 algorithm returns a larger number of predictions, which then needs to be processed and the best ones filtered out. Further adjustments at the field and matrix level are used to select those boxes that are not outside the specified image range or that are not too small or large. In the end, only those boxes that have a sufficiently high prediction success score are returned.

Quality evaluation

The second endpoint API / net / image / quality / is used to evaluate the quality of the goods, ie the collector coin. Like the previous endpoint, this one has the only path to the image. The image is also being loaded. In this case, however, the image is not converted to a field, as the input image will be gradually edited. After loading the coin, the image is converted to grayscale using the built-in OpenCV library function cvtColor(). Subsequently, lines are obtained from the image, which are used to detect circular shapes in the image. To do this, use the HoughCircles() function, which returns the set of all circles found. Using the filter, you will find the circle that represents the outer edges of the coin. This allows us to separate the coin from the outside environment and adjust the background to black, thus removing the redundant information in the image. We will also use gamma correction to highlight the edges and overall relief of the coin. After removing



Fig. 4 Image before and after noise adjustment

the background of the coin, a copy of the image will be created, which will be used to create the perfect version of the coin. The result of this process can be seen in Figure 4.

The next step of the adjustment is thresholding in order to obtain a binary image on which fast and efficient logical operations can be performed.

In this way, two versions of binarized images of the same coin were created, with the difference that one image is free of most damage. Thanks to this, it is possible to proceed to the actual comparison of images and finding the differences. Using the implemented function, the absolute difference of the two image fields is calculated. The resulting field is converted back to a black and white image, creating a difference mask. The contours are then found on this mask, which are then marked in a bold red color. This creates a modified red coin damage mask, which we can add to the original image, thus increasing the space on the coin that shows damage. The result of creating a damage mask and applying this mask to the original image can be seen in Figure 5.



Fig. 5 Original coin and coin with highlighted damage

However, in order to categorize a coin into one of the quality levels, it is necessary to convert this mask into a number that will determine how badly the coin is damaged. The field-level functions add up the pixels that have a defined red color and determine the total number of coins in the coin. Using the ratio of these two data and multiplying by 100, we get the percentage of red pixels on the coin's surface, thus obtaining the desired numerical value of the damage on the coin, which is returned to the server as a result.

EVALUATION

This part of the work will describe the testing and evaluation of the implementation of this portal and its main components, which are used mainly when adding new products. The overall testing will consist of several components:

- usability and user friendliness of the system,
- evaluation of the success of the classification on a larger sample of coins,
- evaluation of categorization on the basis of acquired coin quality.

The testing of the web portal itself was performed with a test sample of 20 users. They were in charge of experimentally using the services offered by the web portal deployed on the production server in terms of usability, user friendliness and learnability of the system. After testing the portal, subjects were asked to fill in a questionnaire created using the methodology of standardized measurement SUS (System usability scale), which, as stated by the author of the article [15] is used to measure the usability and learnability of the system.

The second and third tests are dedicated to the functionality of the web portal in terms of coin type recognition and quality evaluation.

Evaluation of coin type recognition

For this case, a comparison of mAP (Mean Average Precision) values are used, which are also used in testing the accuracy of the trained neural network. Using this value, it is possible to compare the training of the neural network based on the solution of this work with other algorithms. The measurement therefore consists of testing the model on a selected sample of coin images, gradually testing the following cases:

- 1) Images of the right exposure and good quality with the right angle of rotation, bevel of the coin.
- 2) Images of the right exposure and good quality with turned and beveled coins.
- 3) Images with poorer exposure and quality.
- 4) Images with multiple objects and cropped coins.

With the help of these cases, it is possible to obtain a quantitative view of how good a trained neural network is and in which case it starts to show significantly worse results. Testing of each case consists of performing 5 classifications on a sample of 100 randomly selected coin images from the generated datasets. So together we get 20 mAP. As can be seen in Table 1, in the first three cases the model behaved relatively the same and showed similar values as could be seen in Figure 1, where the individual detection algorithms on the COCO dataset were compared. In the fourth case, when it came to images with multiple objects and incorrect exposures, it is possible to see a clear drop in the accuracy of the algorithm. This is precisely because these types of images have been excluded from the dataset, as they do not represent the expected images on the portal.

Table 1
 mAP detection algorithm on tested coin samples for four selected cases

Sample Case /	1. test	2. test	3.test	4.test	5.test
1	55.95%	59.84%	54.56%	61.74%	61.11%
2	58.19%	52.64%	55.66%	57.45%	60.30%
3	53.00%	53.68%	52.79%	54.66%	52.25%
4	30.25%	25.36%	26.86%	25.70%	25.50%

Tests have also shown that some coin grades performed worse than others. This is probably due to the poorer representation of the images of the given classes in the trained dataset. This can be seen in Figure 6, where it is possible to see the average success of the algorithm on the individual tested classes.

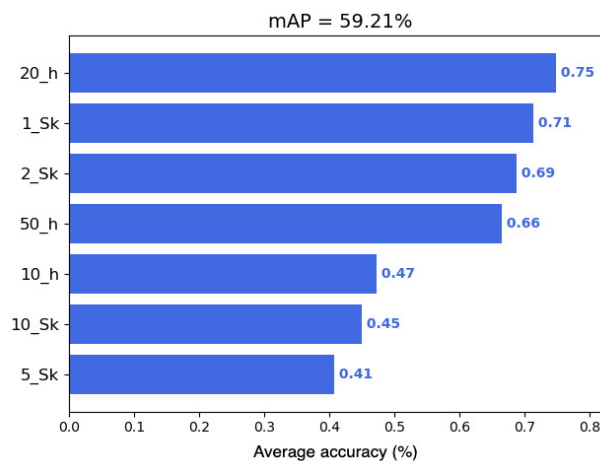


Fig. 6 Average accuracy of the algorithm for individual classes

Evaluation of quality determination

As there is no standardized, common procedure for this case, it is necessary to create your own measurement, which will allow quantitative assessment of whether the coin quality assessment process behaves correctly and categorizes coins to the correct quality levels. web portal 1 which are already categorized into quality levels. Using this data, it will be possible to compare whether the quality evaluation process implemented in this solution provides correct results and thus whether

the coins will be correctly evaluated and the assigned quality levels will match the levels on the numismatic portal. This testing will be performed on 140 coins from this portal, using images of the coins on the portal for testing purposes and used to test the implementation of the algorithm of this work. These results will then be converted into a binary format expressing whether or not the quality level we determine matches the level on the numismatic portal. Subsequently, the percentage of records that were determined correctly is determined, which quantifies the success of our algorithm. Table 2 shows the results of quality assurance testing on individual coin types. It is clear from the displayed results that the overall success of the algorithm was higher than 50%, which means that every second coin was correctly evaluated and assigned the correct quality level.

Table 2
Success of quality evaluation on types of coins

Coin	Total number	Correctly determined	Success
10 h	20	12	60%
20 h	20	17	85%
50 h	20	11	55%
1 Sk	20	9	45%
2 Sk	20	13	65%
5 Sk	20	15	75%
10 Sk	20	7	35%
Total	140	84	60%

The worst results are observable with 10 Sk and 1 Sk coins. This may be due to the color of the material, as these coins of all tested are the only gold in color that could change to too dark when the color scale changed to gray, which would lose important coin features. Another reason for the low success rate for these types of coins may be the fact that these coins have the most complex designs with different notches, which could have been falsely identified as damage to the coin. These misidentified damages are probably caused by poor thresholding and binarized image creation. The solution in this case could be a better setting of the specified attributes, which could depend on the complexity of the relief of the coin. From the above results of the coin quality assessment, it is also possible to see the opposite trend as described so far for the 20 h and 5 Sk coins. In this case, too, the complexity of the coin design could be a possible explanation, as these coins, on the other hand, probably have the finest and clearer edges. However, this possibility did not always prove to be true, as the edges of these coins were also incorrectly assessed as damage. The reason may be the subtlety of the edges, which were removed along with the shortcomings in the refinement of the noise in making a perfect copy of the coin. After the correction of this process, a change in the quality assessment was recorded, which was also reflected in the correct determination of the quality level. After the mentioned modifications, the same test was performed on the same images again. The results are shown in Table 3. It is possible to see an overall improvement of more than 12%, which results in particular in improving the success of quality assessments for 1 Sk and 10 Sk coins.

Table 3
Success of quality evaluation on types of coins after changes

Coin	Total number	Correctly determined	Success
10 h	20	15	75%
20 h	20	19	95%
50 h	20	13	65%
1 Sk	20	14	70%
2 Sk	20	13	65%
5 Sk	20	17	85%
10 Sk	20	10	50%
Total	140	101	72.14%

CONCLUSION

The purpose of this paper was to create a web portal that would offer users the opportunity to add their own goods and find out its type and quality. The aim was to use modern techniques of object detection in the image, thanks to which it would be possible to classify the type of object that would be displayed to users of the portal. At the beginning of this work, the most used algorithms for detecting objects in the image were analyzed. Their basic architecture, their advantages and disadvantages were described. Finally, these algorithms were compared and the YOLO v3 algorithm was selected, which recorded the highest detection accuracy with the

relatively fastest response. Another part of the analysis was devoted to determining the quality of goods - collector coins. As this is a specific type of goods, it was first necessary to find out how the quality of the coins is evaluated and what features and features of the coin are important. Subsequently, it was necessary to ensure that the quality of the coin images was sufficient to assess the quality on the basis of this knowledge. Based on the knowledge gained from the analysis of the correct design of the coin images, the criteria that had to be met for the input coin image from the user were determined.

With the help of the created design, it was possible to proceed to the actual implementation of the web portal. This section describes the architecture of the system, its external components and how to communicate with them, but also the technologies and libraries used. In the last chapter of this work, an evaluation was presented, which consisted of three types of testing, which were on the one hand focused on the user experience and usability and clarity of the web portal, but on the other hand on the functional side of the system. Using standardized testing and comparing with the existing results of other portals, the success of coin type prediction and coin quality evaluation was quantified. From the obtained results, possible causes of system error were interpreted and possible solutions were proposed, which were subsequently implemented and retested. The implemented solution offers several possibilities for improvement. One of them is, in particular, the improvement of the algorithm for evaluating the quality of collector coins, which would offer the user clearer information about what specific damages and deficiencies are found on his added coin. Another possible improvement would be modifications to the client part, which would mean better usability of the web portal. These changes would mainly involve faster page response and overall process optimization.

ACKNOWLEDGMENT

This publication is the result of the Project implementation: University Science Park TECHNICOM for Innovation Applications Supported by Knowledge Technology, Phase II., ITMS: 313011D232, supported by the Research Inovation Operational Programme funded by the ERDF. We support research activities in Slovakia/This project is being co-financed by the European Union.

REFERENCES

- [1] GIRSHICK, Ross; DONAHUE, Jeff; DARRELL, Trevor; MALIK, Jitendra. Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2014, s. 580–587.
- [2] REN, Shaoqing; HE, Kaiming; GIRSHICK, Ross; SUN, Jian. Faster r-cnn: Towards real-time object detection with region proposal networks. arXiv preprint arXiv:1506.01497. 2015.
- [3] LIU, Wei; ANGUELOV, Dragomir; ERHAN, Dumitru; SZEGEDY, Christian; REED, Scott; FU, Cheng-Yang; BERG, Alexander C. Ssd: Single shot multibox detector. In: European conference on computer vision. 2016, s. 21–37.
- [4] REDMON, Joseph; DIVVALA, Santosh; GIRSHICK, Ross; FARHADI, Ali. You only look once: Unified, real-time object detection. Proceedings of the IEEE conference on computer vision and pattern recognition. 2016, s. 779–788.
- [5] REDMON, Joseph; FARHADI, Ali. Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767. 2018.
- [6] NGC Coin Grading Scale [https://www.ngccoin.com/coin-grading/grading-scale/]. [B. r.]. Navštívené: 2021-09-30.
- [7] HOBERMANN, Gerald. The art of coins and their photography: an illustrated photographic treatise with an introduction to numismatics. 1981.
- [8] KAMPEL, Sebastian ZAMBANINI–Martin. Coin data acquisition for image recognition. In: Proceedings of the 36th CAA Conference. 2008, zv. 2, s. 6.
- [9] PATRO, S; SAHU, Kishore Kumar. Normalization: A preprocessing stage. arXiv preprint arXiv:1503.06462. 2015.
- [10] ATIGHEHCHIAN, Parmida. Coin wear estimation and automatic coin grading. 2017. Diz. pr. Concordia University.
- [11] HUA, Xia. Human computer interactions for converting color images to gray. Neurocomputing. 2012, roč. 85, s. 1–5.
- [12] GÜNEŞ, Ali; KALKAN, Habil; DURMUŞ, Efan. Optimizing the Color-toGrayscale Conversion for Image Classification. Signal Image and Video Processing. 2016. Dostupné z doi: 10.1007/s11760-015-0828-7.
- [13] MODI, Shatrughan; BAWA, Dr et al. Automated coin recognition system using ANN. arXiv preprint arXiv:1312.6615. 2013.
- [14] GARETH, James; DANIELA, Witten; TREVOR, Hastie; ROBERT, Tibshirani. An introduction to statistical learning: with applications in R. Spinger, 2013.
- [15] SAURO, Jeff. Measuring usability with the system usability scale (SUS). 2011. Dostupné tiež z: https://measuringu.com/sus/.

Porovnanie Návrhov Fotovoltaiky na Rodinný Dom

¹Marek PAVLÍK, ²Dávid OLŠINÁR

^{1,2} Department of Electric Power Engineering, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Slovak Republic

¹marek.pavlik@tuke.sk, ²david.olsinar@student.tuke.sk

Abstract — Tento príspevok sa zaoberá problematikou fotovoltaických systémov. V článku sú rozobrané dva prípady. Oba prípady majú rovnaké parametre fotovoltaiky avšak jeden prípad je s orientáciou na sever a druhý na juh. Prípady boli simulované v programe PV-SOL. V tomto prípade bol potvrdený predpoklad, že orientácia fotovoltaiky na juh je priaznivejšia a ekonomicky výhodnejšia ako orientácia na sever. Potvrdili to aj simulácie prezentované v tomto príspevku.

Keywords — fotovoltaický článok, MPP bod, napätie naprázdno, prúd nakrátko, string

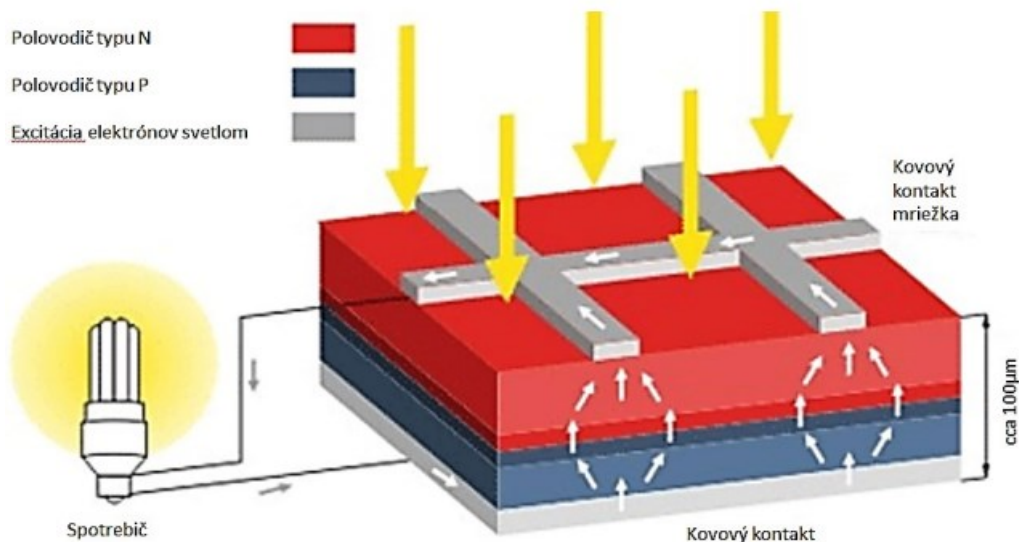
I. ÚVOD

V posledných rokoch neustále rastie dopyt po technológiách, ktoré sa zaoberajú využitím alternatívnych zdrojov energií, no aj napriek tomu na Slovensku nie sú tieto zdroje energií až tak populárne. Hlavným dôvodom je nízka miera informovanosti obyvateľov Slovenskej republiky k danej problematike, no taktiež môže ísť o strach z niečoho nového. Ako ďalším dôvodom by mohla byť nedôvera investovania do daných alternatívnych zdrojov elektrickej energie, najmä z pohľadu ceny, hodnovernosti, účinnosti a takisto z pohľadu návratnosti investícií. [1] [2]

Energia získaná zo Slnka patrí medzi primárne zdroje na Zemi, vďaka nej sa vytvorili niektoré základné podmienky pre život. Táto energia pochádzajúca zo Slnka, ktoré spôsobuje vznik zrážok, vyparovanie vody, poveternostné podmienky, rast, fotosyntézu flóry a podobne. S gravitačnou silou Zeme pôsobí na celý kolobeh vodných tokov v prírode. Tým pádom je pôvodcom väčšiny bežne dosiahnuteľných energií na Zemi s výnimkou teploty zemského jadra a energiou morí a oceánov. [2]

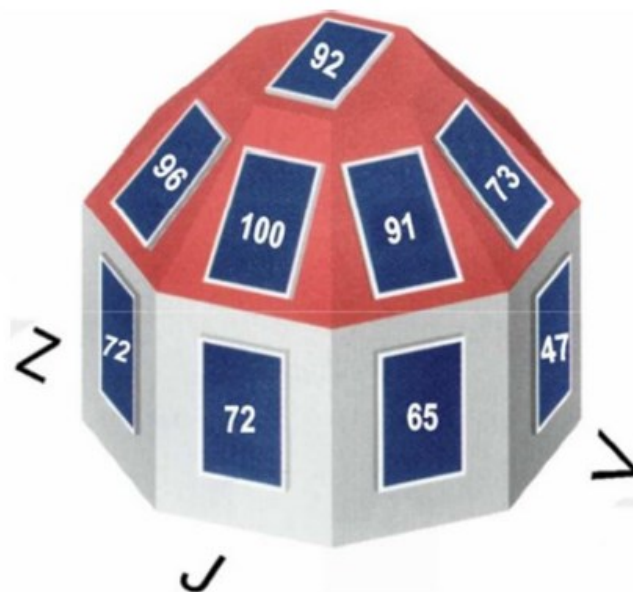
Slnčná energia má viacero spôsobov využitia, kde najefektívnejším spôsobom je využitie fototermálnej premeny slnečného žiarenia pre zabezpečenie akumulovaného tepla. Poznáme taktiež fotovoltaickú premenu lenže tá je menej efektívna a sprevádzajú ju vyššie náklady.

Technológia využitia slnečnej energie v podobe fotovoltaických článkov funguje na základe toku elektrického prúdu medzi vzájomne prepojenými polovodičmi, ktoré majú rozdielne elektrické vlastnosti. V jednej vrstve kremíka (Obr.1) je prebytok elektrónov za čo môže prísť atómov fosforu, táto vrstva sa nazýva N-vrstva. V druhej vrstve kremíka je nedostatok elektrónov za čo môže prísť atómov bóru, táto vrstva sa nazýva P-vrstva. Medzi týmito vrstvami sa nachádza P-N prechod. Ten sa aktivuje za pomoci slnečného žiarenia medzi oboma vrstvami a začína medzi nimi tečť elektrický prúd. Výkon týchto systémov závisí od intenzity slnečného žiarenia v danej lokalite. [1] [2]



Obr. 1 Princíp činnosti fotovoltaického článku

Na účinnosti a celkovom zisku slnečnej energie sa podieľa ako orientácia tak aj sklon FV panelov. Z obr. 2 vyplýva, že ideálna orientácia je smerom na juh alebo juhozápad, na Slovensku je priaznivý sklon 30 až 35°. [1] [2]



Obr. 2 Zobrazenie účinnosti FV panela od orientácie a sklonu

II. SPÔSOBY ZAPOJENIA FOTOVOLTAICKÝCH SYSTÉMOV

A) Systém On-grid

Tieto systémy sú pripojené priamo na verejnú sieť, dodávajú do nej elektrickú alebo v prípade nevyhnutnej potreby z nej odoberajú elektrickú energiu. Schéma zapojenia On-grid systému je znázornená na obr.3.

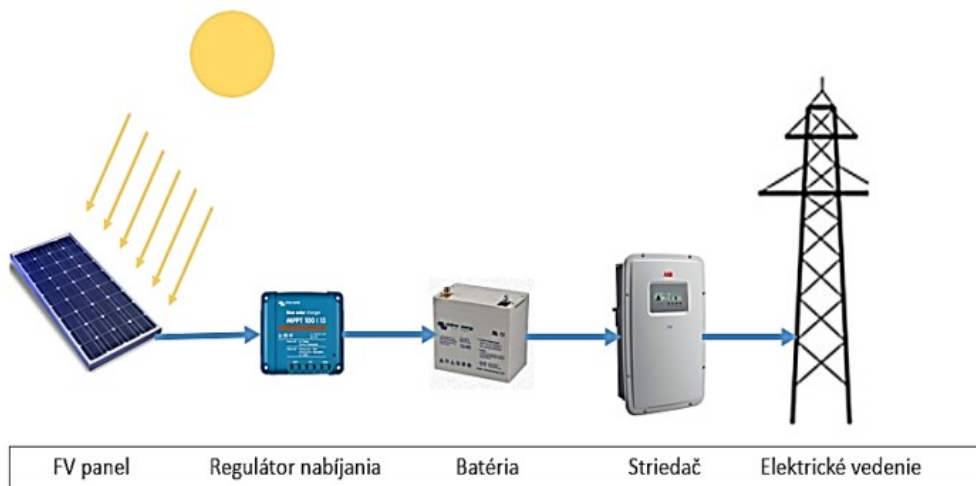
B) Ostrovný systém (Of-grid)

Tieto systémy nie sú napojené na verejnú sieť. Ich úlohou je akumulácia energie do batérii v čase, kedy nesvieti Slnko. Jednou z častí tohto systému je aj kontrolný mechanizmus, ktorý zabezpečuje aby nedošlo k zbytočnému nabíjaniu a vybíjaniu akumulátora. Schéma zapojenia Off-grid systému je znázornená na obr.4.

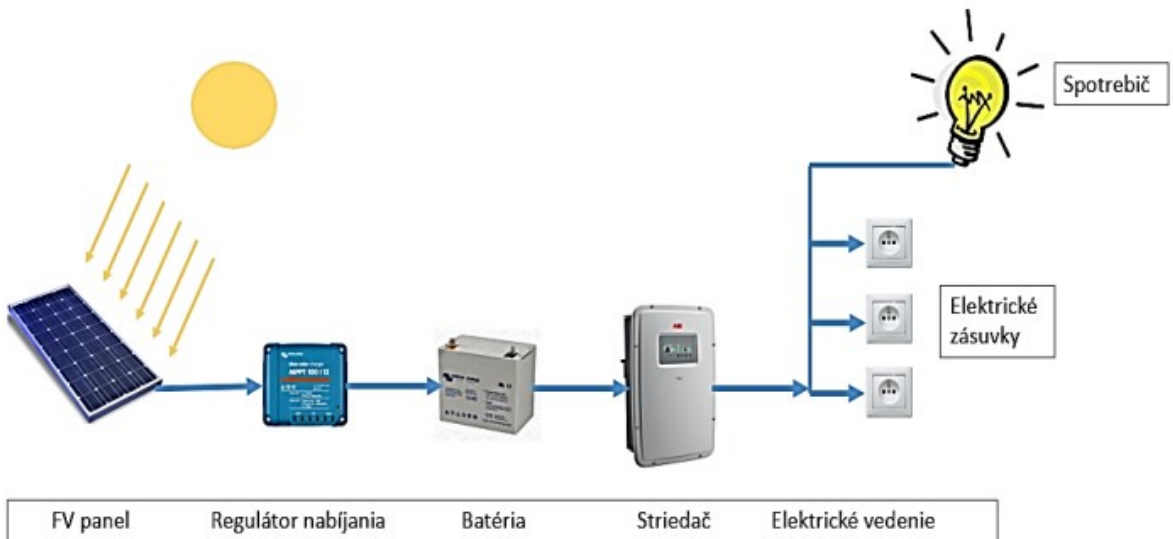
C) Hybridný systém

Tieto systémy zvyknú spolupracovať s inými zdrojmi energie, ako napríklad biomasa, veterná turbína, diesellový generátor a pod. Tým že je to kombinovaný systém, je možné ho pripojiť do

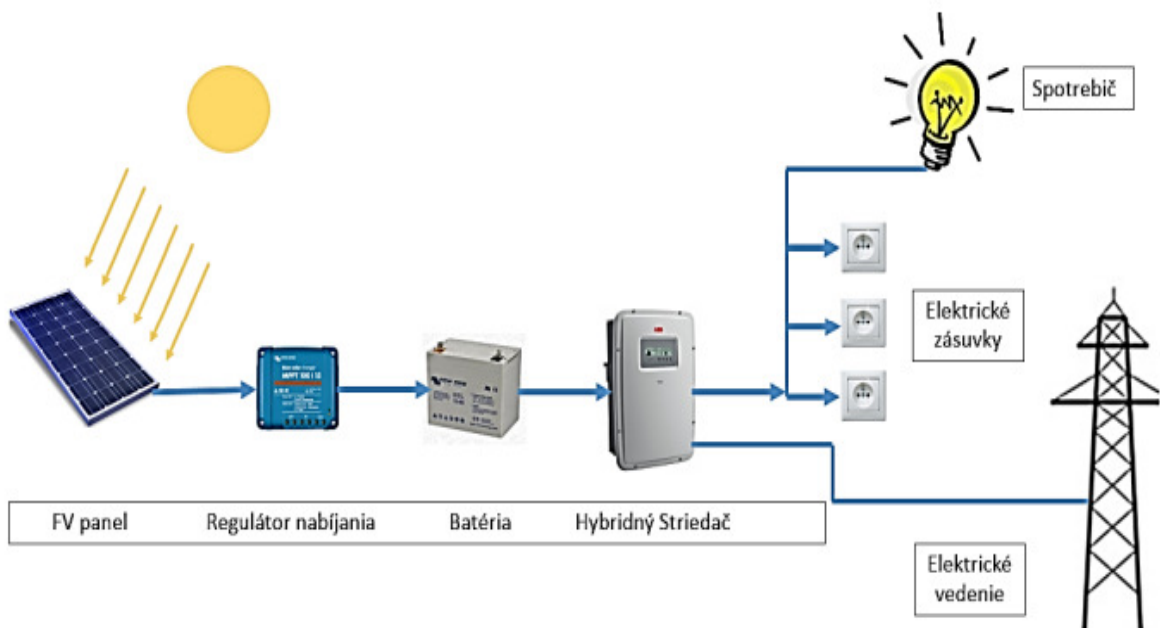
siete alebo samostatne. Prevažne sú využívané tieto systémy pre menšie budovy a rodinné domy. Schéma zapojenia Hybridného systému je znázornená na obr. 5.



Obr. 3 Systém On-grid



Obr. 4 Systém Off-grid



Obr.5 Hybridný systém

III. ANALÝZA NÁVRHOV FOTOVOLTAICKÝCH SYSTÉMOV PRE RODINNÝ DOM

V danom návrhu pre rodinný dom uvažujeme s hybridným systémom, ktorého priemerná ročná spotreba elektrickej energie predstavuje 4308 kWh za rok, táto hodnota je nameraná pre priemernú rodinu s 2 osobami a 2 deťmi. Typ strechy je sedlový, ktorý je orientovaný na sever a juh. Strecha má 30° uhol, nakoľko na území Slovenskej republiky je najvhodnejší sklon pre fotovoltaické panely 30-35° z dôvodu vyššej účinnosti. Panely využívané pri návrhu sú monokryštalické, nakoľko tieto sú najviac používané. Otázka, ktorou je potrebné sa zaoberať je orientácia fotovoltaických panelov smerom na sever alebo juh a takisto vyhodnotenie účinnosti fotovoltaických panelov bez zatienenia alebo so zatienením, pretože ako je známe pri monokryštalických fotovoltaických paneloch účinnosť prudko klesá už pri miernom zatienení. Práve kvôli tomu boli vytvorené viaceré simulácie v programe PV*SOL premium, za účelom analýzy a záverečného vyhodnotenia. [1] [3]

Pre dané návrhy boli vybrané fotovoltaické panely od firmy Swisswatt a to konkrétne typu MW200 o počte 30 kusov – technické parametre vid' Tab.1. Panely Swisswatt MW200 sú monokryštalické a ich rozloha predstavuje 38,3 m² s celkovým inštalovaným výkonom 6 kW. Typ montáže je paralelný so strechou a taktiež sú panely dobre odvetrávané zo zadnej strany. [3] [4]

Tab. 1 Technické parametre fotovoltaického panela

Swisswatt MW200	Parametre
Typ	Monokryštal
Počet	30 kusov
Špičkový výkon	200 Wp
Menovité napätie	37,3 V
Menovitý prúd	5,43 A
Napätie naprázdno	45,7 V
Skratový prúd	5,72 A
Napätie pri maximálnom výkone	33,5 V
Prúd pri maximálnom výkone	4,52 A
Účinnosť panelu	15,66%
Rozsah prevádzkových teplôt	-40 až 85 °C
Teplotný koeficient Pmax	0,4454 %/°C
Teplotný koeficient Voc	0,3336 %/°C
Teplotný koeficient Isc	0,0499 %/°C
Menovitá pracovná teplota	47 ± 2 °C
Maximálne napätie systému	1 000 V
Rozmer panelu (VxŠxH)	1 580 x 880 x 40 mm
Rozmer bunky	125 x 125 mm
Hrúbka skla	3,2 mm
Stupeň krytia IP	IP 67
Hmotnosť panela	15,5 kg
Záruka na produkt	10 rokov
Záruka napájania (degradácia)	12 rokov 90% vstupného výkonu, 25 rokov 80% vstupného výkonu

Pre dané návrhy bol vybraný menič od firmy Huawei Technologies a to konkrétne typu SUN2000-6KTL-MO. Menič pozostáva z 2 MPP vstupov, kde do každého zo vstupov je sériovo napojených 15 fotovoltaických modulov. Na menič je priamo napojený batériový systém. Ďalej, pre dané návrhy bol vybraný akumulátor od firmy Huawei Technologies a to konkrétne typu LUNA2000-10-SO. Zväčša akumulátor býva najdrahšou položkou pre hybridný systém. Daný batériový systém pozostáva z dvoch akumulátorov, kde kapacita jedného je 5 kWh. Technológia akumulátora je Lítium-železo-fosfát (LiFePo4). Tento systém je priamo napojený na menič, preto je možné nabíjanie akumulátora iba prostredníctvom fotovoltaických panelov. [1] [5]

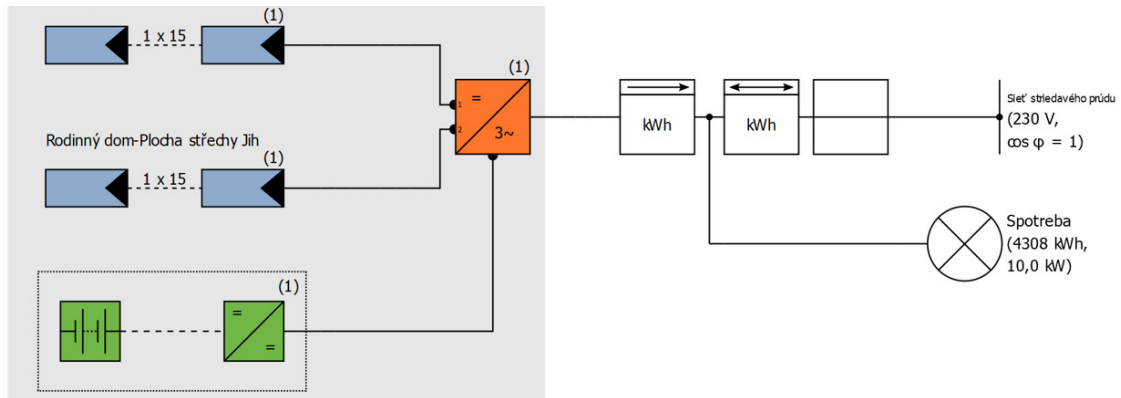
Tento systém dodáva elektrickú energiu do rodinného domu pre vlastnú spotrebu domácnosti, v prípade prebytku je elektrická energia dodávaná do verejnej siete, prípadne ak je elektrickej energie nedostatok, zvykne byť odoberaná z verejnej siete za účelom pokrytia vlastnej spotreby.

Na obr. 6 sa nachádzajú fotovoltaické panely o počte 30 kusov, ktoré sú rozdelené na 2 vetvy, kde v jednej vetve sa nachádza 15 kusov fotovoltaických panelov, tieto vetvy sú napojené do 2 MPP vstupov meniča. Batériový systém je priamo napojený na menič. [1]

Menič v prípade prebytku elektrickej energie v batériovom systéme dodáva túto energiu buď pre vlastnú spotrebu domácnosti alebo do siete pokiaľ nie je potrebné zásobovať rodinný dom.

V domovej prípojke sa nachádza feed-in elektromer a obojsmerné počítadlo, nakoľko sa jedná o hybridný systém. [1]

Pod schematickým zobrazeným daného fotovoltaického systému sa nachádza legenda, v ktorej sú zhrnuté všetky zariadenia danej schémy, popisuje konkrétne typy fotovoltaických panelov, meniča a batériového systému, taktiež poukazuje na niektoré z parametrov. [1]



Obr.6 Schematické zapojenie FV systému

V tejto časti boli porovnávané dva spôsoby orientácie fotovoltaických panelov a to konkrétne na sever a na juh. Z tab. 2 je možné zhodnotiť že špecifický ročný výnos pri severnej orientácii predstavuje hodnotu 611,08 kWh/kWp čo je skoro polovica z toho čo vyrobí fotovoltaický systém s orientáciou na juh, kde táto hodnota je 1 134,87 kWh/kWp. Taktiež južná orientácia má väčší stupeň využitia zariadenia približne o 9%. Energetický výnos je priaznivejší z pohľadu podielu vlastnej spotreby pre severnú orientáciu, nakoľko väčšia časť elektrickej energie je využitá pre pokrytie spotreby rodinného domu a to 2 251 kWh/rok a do siete dodáva 1 164 kWh/rok, južná orientácia naopak dodáva o niečo viac elektrickej energie do siete ako pre vlastnú spotrebu. Hodnota pre vlastnú spotrebu predstavuje 3 152 kWh/rok a hodnota elektrickej energie dodanej do siete je 3 358 kWh/rok. Energia dodávaná do siete zabezpečuje väčšie príjmy z predanej elektrickej energie. [1]

Celková spotreba je takmer identická, no pokrývanie pomocou fotovoltaického systému s batériou je opäť výhodnejšie pre južnú orientáciu, kde predstavuje hodnotu 3 152 kWh/rok, pre severnú orientáciu je táto hodnota 2 251 kWh/rok. Pri severnej orientácii je menej elektrickej energie dodávanej cez fotovoltaické panely s batériou, preto je potrebné viac elektrickej energie kryť pomocou siete ako pri južnej orientácii, tým pádom je podiel pokrytia zo strany solárnej energie priaznivejší pre juh s podielom 73% ako pre sever kde tento podiel je 52%.

Systém batérií má vyššiu životnosť pri severnej orientácii a to výše 20 rokov, nakoľko je menej zaťažovaný ako to je pri južnej orientácii, kde životnosť predstavuje 15 rokov, preto je potrebné počítať s nutnosťou dokúpenia nového akumulátora v prípade nefunkčnosti predošlého, ktorého životnosť sa skončí približne v 15 roku fungovania systému. [1]

Tab. 2 Analýza hospodárnosti prevádzky FV systému [1]

	Sever	Juh
Fotovoltaické zariadenie		
Celkový výkon fotovoltaických článkov	6 kWp	
Špecifický ročný výnos	611,08 kWh/kWp	1 134,87 kWh/kWp
Koeficient využitia zariadenia	76,7%	84,3%
Zníženie výnosov z dôvodu zatienu	0%	
Energetický výnos (AC sieť)	3 481 kWh/rok	6 510 kWh/rok
Priama vlastná spotreba	2 251 kWh/rok	3 152 kWh/rok

Napájanie do siete	1 230 kWh/rok	3 358 kWh/rok
Podiel vlastnej spotreby	64,5%	48,4%
Spotrebiteľ		
Spotrebiteľ	4 308 kWh/rok	
Spotreba v pohotovostnom režime (Menič)	16 kWh/rok	8 kWh/rok
Celková spotreba	4 324 kWh/rok	4 316 kWh/rok
Krytie spotreby pomocou FV a batérie	2 251 kWh/rok	3 152 kWh/rok
Krytie spotreby pomocou siete	2073 kWh/rok	1 164 kWh/rok
Podiel pokrytia solárnou energiou	52,10%	73,00%
Systém batérií		
Nabíjanie na začiatku	10 kWh	
Nabitie batérie pomocou FV	1 756 kWh/rok	2 598 kWh/rok
Výkon batérie pre pokrytie spotreby	1 554 kWh/rok	2 291 kWh/rok
Straty nabíjaním/vybíjaním	192 kWh/rok	277 kWh/rok
Straty v batérii	19 kWh/rok	40 kWh/rok
Zaťaženie cyklov	4,40%	6,50%
Životnosť	>20 rokov	15 rokov

V Tab. 3 je možné vidieť náklady na výrobu elektrického prúdu sú menšie pre južnú orientáciu približne o polovicu. Zisk fotovoltaického systému orientovaného na juh v prvom roku predstavujú hodnotu viac ako dvojnásobnú v porovnaní so severom. Úspory v prvom roku sú taktiež priaznivé pre juh kde suma 616,71€ je vyššia ako 445,49€ v prípade severu. Aj v tomto prípade je juh výhodnejší.

Tab. 3 Analýza hospodárnosti fotovoltaických systémov [1]

	Sever	Juh
Sietové napájanie v prvom roku (vrátane degradácie modulu)	1 230 kWh/rok	3 313 kWh/rok
Sledované obdobie	20 rokov	
Úrok z kapitálu	1%	
Ekonomické parametre		
Interná miera návratnosti (IRR)	2,65%	6,08%
Kumulatívny cashflow	1 820,29 €	5 553,02 €
Doba návratnosti	17,6 roka	12,2 roka
Náklady na výrobu prúdu	0,1377 €/kWh	0,0734€/kWh
Prehľad platieb		
Špecifické investičné náklady	1500 €/kWp	
Investičné náklady	9 000 €	
Revízia a úspory		
Celkové provízie v prvom roku	63,12 €/rok	170,06 €/rok
Úspory v prvom roku	445,49 €/rok	616,71 €/rok
Výkupný tarif		
Špecifická provízia z napájania	0,0513 €/kWh	
Výkupný tarif	63,1216 €/rok	170,0577 €/rok
Distribučná sadzba		
Pracovná cena	0,1993 €/kWh	
Základná cena	2,88 €/mesiac	
Koeficient ceny elektrickej energie	2%/rok	

IV. ZÁVER

V tomto príspevku sa autori zamerali na problematiku fotovoltaických systémov. Tieto systémy sú v súčasnosti veľmi využívané aj vďaka podpory, ktoré je momentálne pozastavená. Keďže medzi výhody slnečných systémov patrí ich nevyčerpatelnosť, využitie slnečnej energie patrí k jednému z najzaujímavejších a dostupných zdrojov energie pre domácnosť.

Tento príspevok poukazuje na fakt, že jedným z dôležitých parametrov, ktoré ovplyvňujú

výrobu elektrickej energie z fotovoltaiiky je orientácia. Bol predpoklad, že orientácia na juh je pravdepodobne najefektívnejšia. Tento fakt sa simuláciami v tomto príspevku potvrdil. Náklady na výrobu elektrického prúdu sú menšie pre južnú orientáciu približne o polovicu. Zisk fotovoltaiického systému orientovaného na juh v prvom roku predstavujú hodnotu viac ako dvojnásobnú v porovnaní so severom. Úspory v prvom roku sú taktiež priaznivé pre juh kde suma 616,71€ je vyššia ako 445,49€ v prípade severu. Aj v tomto prípade je juh výhodnejší.

POĎAKOVANIE

Táto práca bola podporená Agentúrou na podporu výskumu a vývoja na základe zmluvy č. APVV-19-0576, ako aj Vedeckou grantovou agentúrou Ministerstva školstva, vedy, výskumu a športu SR a SAV, podporená grantom VEGA 1/0757/21.

LITERATÚRA

- [1] D. Olšinár, „Porovnanie rôznych návrhov fotovoltaiických systémov pre rodinný dom,“ Bakalárska práca, 2022.
- [2] M. Pavlík, „Obnoviteľné zdroje energie vo všeobecnosti,“ 1.vyd. Technická univerzita v Košiciach, 2019. 75 s. ISBN 978-80-553-3317-5.
- [3] J. Zhang, et. al., "Model Parameter Analysis of Cracked Photovoltaic Module under Outdoor Conditions," 2020 47th IEEE Photovoltaic Specialists Conference (PVSC), 2020, pp. 2509-2512, doi: 10.1109/PVSC45281.2020.9300720.
- [4] N. Das, et.al., "Modelling and analysis of multi-junction solar cells to improve the conversion efficiency of photovoltaic systems," 2014 Australasian Universities Power Engineering Conference (AUPEC), 2014, pp. 1-5, doi: 10.1109/AUPEC.2014.6966482.
- [5] T. Maniak, "Modelovanie parametrov fotovoltaiických článkov," bakalárska práca, 2021, Technická univerzita v Košiciach, pp.77.

Návrh netypizovaného energetického zdroja a jeho praktické využitie

Martin BINAR, Dušan MEDVEĎ

Katedra elektroenergetiky, Fakulta elektrotechniky a informatiky,
Technická univerzita v Košiciach, Slovenská republika

`martin.binar@student.tuke.sk, dusan.medved@tuke.sk`

Abstrakt — Tento príspevok sa zaoberá návrhom netypizovaného energetického zdroja a jeho následné praktické využitie. Ako netypizovaný zdroj bol zvolený batériový modul, ktorý je napájaný fotovoltaickým článkom. Pri návrhu sa uvažovali jednotlivé nelineárne charakteristiky použitých prvkov, ktoré sa následne využili pri vyčíslení celkovej energetickej bilancie energetického zdroja. Riadiacim prvkom bol zvolený modul Arduino Mega2560, pretože má dostatočný počet I/O konektorov. Keďže nelineárne charakteristiky pri použití kombinácie viacerých prvkov sú ovplyvňované viacerými faktormi, preto bola realizovaná doplňujúca termovízna kontrola pre zistenie skutočného stavu oteplenia. V závere príspevku sú uvedené výsledky merania na zhotovenom modeli energetického zdroja.

Kľúčové slová — fotovoltaický panel, nekonvenčný zdroj energie, účinnosť, energetická bilancia, batéria

I. ÚVOD

Pre účely tohto príspevku sme sa zamerali na vytvorenie zdroja elektrickej energie, ktorý je čistý, môže sa využívať každý deň, či už v domácnosti alebo mimo nej. Preto som sme si vybrali ako zdroj fotovoltaický panel, ktorý bude konvertovať energiu zo slnka na elektrickú energiu. Súčasný fotovoltaický panely dokážu pracovať s účinnosťou okolo 15÷17 %. To znamená, že z celkovej energie slnka, ktorú prijmu, len 15÷17 % vedia premeniť na elektrickú energiu.

V prípade, ak tento fotovoltaický panel nebude vedieť vyrobiť dostatok elektrickej energie, bude sa používať batéria ako záložný elektrický zdroj. Vďaka tomu sa dosiahne stálosť elektrického napätia na výstupe aj pri zhoršenom počasi, prípadne napájania zariadenia aj v nočných hodinách. Prebytočná energia sa bude využívať na dobíjanie batérie. Takto dokážeme využiť čo najviac elektrickej energie vyrobenej z fotovoltaického panela a uchovať ju pre neskoršie využitie.

II. NÁVRH A IMPLEMENTÁCIA RIEŠENIA NETYPIZOVANÉHO ENERGETICKÉHO ZDROJA

Pri zhotovovaní tohto netypizovaného energetického zdroja bolo uvažované tak, aby tento zdroj elektrickej energie bol jednoduchý na zostrojenie, bol finančne dostupný a mal praktické využitie v bežnom každodennom používaní.

Pod pojmom netypizovaný energetický zdroj energie si môžeme predstaviť zdroj energie, ktorý sa bežne nevyužíva v domácnosti.

A. Konštrukcia zdroja elektrickej energie

Keďže tento netypizovaný elektrický zdroj tvorí fotovoltaický (FV) panel, ako zdroj primárnej elektrickej energie, bolo nutné zvoliť vhodný fotovoltaický panel, ktorý bude vyhovovať stanoveným požiadavkám.

Pri konštrukcii netypizovaného elektrického zdroja bolo nutné počítať s rôznymi podmienkami, ktoré boli vopred stanovené:

- adekvátne investičné náklady,
- jednoduchosť zostavenia konštrukcie,
- použitie vhodných súčiastok.

Presný finančný limit nebol stanovený. Bolo však potrebné určiť vybrané periférie zariadenia tak, aby bolo možné ich vymeniť bez väčších finančných nárokov, ako sú napríklad výmena fotovoltaického panela s vyšším výkonom alebo s väčšou účinnosťou, výmena akumulátorov s väčšou kapacitou.

Konstruované zariadenie je možné rozdeliť na niekoľko častí:

- fotovoltický panel,
- akumulátor,
- doska meničov,
- doska merania a spracovania údajov,
- dátovo komunikačné zariadenia.

1) Fotovoltický panel

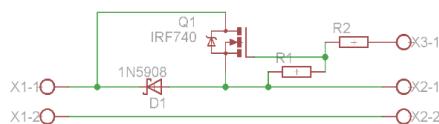
Ako fotovoltický panel bol vybraný cenovo dostupný fotovoltický panel s adekvátnym výkonom, aby nebolo nutné sa veľmi obmedzovať pri výbere zariadenia, ktoré je možné napájať. Pri výbere fotovoltického panela s väčším výkonom, dokáže celé zariadenie napájať všetky dostupné zariadenia, ktoré budú v danej energetickej úrovni. Dôležitým parametrom je aj to, že daný panel nemusí vždy vzhľadom na počasie produkovať maximálny výkon. Práve preto bolo rozhodnuté použiť monokryštalický fotovoltický panel „Maxx 30W“ od výrobcu SOLAR. V ideálnych podmienkach a maximálnom výkone panela je na výstupe meniča možné získať prúd až 2,5 A. Táto hodnota prúdu umožňuje napájať aj väčšie spotrebiče, či nabíjať mobilný telefón pomocou technológie Fast Charging, resp. rýchlo-nabíjania.

2) Akumulátor

V prípade prebytku energie sa táto energia uskladní v batérii akumulátorov. Okrem iného, pomocný akumulátor bude použitý aj pre prípad nedostatočného, resp. žiadneho výkonu solárneho panela na napájanie spotrebiča.

Z dôvodu jednoduchosti použitia a dobrej dostupnosti, boli použité batérie typu Li-Ion o veľkosti 18650. Tieto boli vyskladané do sério-paralelného zapojenia tak, aby bolo možné dosiahnuť požadovanú kapacitu a prevádzkové napätie. Jeden takýto článok má minimálnu hodnotu napätia 2,5 V a špičkovú hodnotu 4,2 V. Menovité napätie Li-Ion 18650 akumulátorov je 3,7 V. Tento typ akumulátora dosahuje prevádzkovú kapacitu až 3500 mAh.

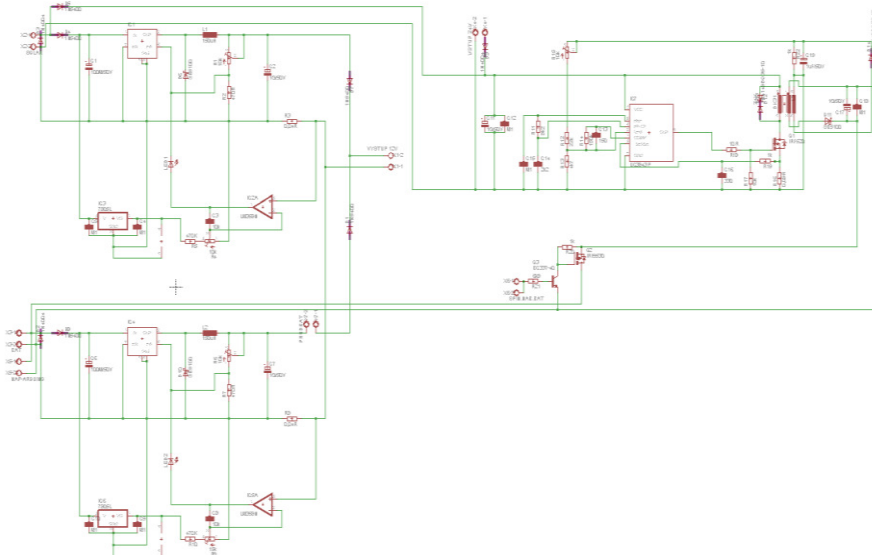
Batéria akumulátorov je chránená aj ochranou pred podbitím. Tá je realizovaná digitálne pomocou procesora. Pomocou merania napätia akumulátora procesor vyhodnotí, či celkové napätie batérie nekleslo pod stanovenú hodnotu. Pokiaľ sa tak stalo, tak odpojí napätie z riadiacej elektródy spínacieho tranzistora a záťaž sa odpojí. Nabíjanie akumulátora je zaručené antiparalelnou diódou.



Obr. 1 Schéma zapojenia ochrany pred podbitím batérie

3) Doska meničov

Doska meničov zabezpečuje tri základné úlohy: napájanie spotrebiča pomocou solárneho panela, napájanie spotrebiča pomocou akumulátora a nabíjanie batérie.



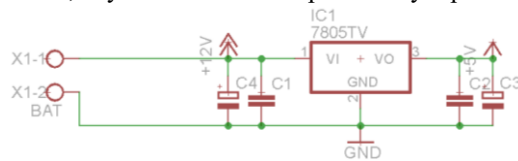
Obr. 2 Schéma zapojenia dosky meničov

4) Meracia a riadiaca doska plošného spoja

Táto doska obsahuje mikrokontrolér Arduino Mega2560, na ktorý sú napojené meracie zariadenia prúdu a napätia. Toto Arduino bude vykonávať merania a následne sa tieto hodnoty budú zobrazovať na displeji. Napájanie bude realizované zo stabilizátora napätia na 5 V a napájané bude priamo z akumulátora, aby bol zabezpečený jeho konštantný chod. Keďže hlavnou úlohou je vyhodnotenie energetickej bilancie, je potrebné zbierať dáta zo vstupov a výstupov v obvode.

Pre snímanie napätia sa používal senzor napätia do 25 V VST1401 a pre snímanie prúdu sa používal senzor prúdu do 5 A ACS712.

Prúd a napätie sa snímali na vstupe do silovej dosky zo solárneho panela, resp. z výstupu zo solárneho panela. Taktiež sa bude merať napätie a prúd na výstupe. Meranie napätia na akumulátore bude hlavne na zistenie stavu nabitia akumulátora, čiže zistenie percentuálneho stavu nabitia batérie, kedy hodnota pod 12 V bude 0 %, nakoľko s takou hodnotou už nie je možné napájať obvod na 12 V. Tento snímač slúžil aj ako senzor podbitia, ktorý odpojil batériu od obvodu pri určitej hodnote, aby sa akumulátor nepodbil a tým pádom nezničil.

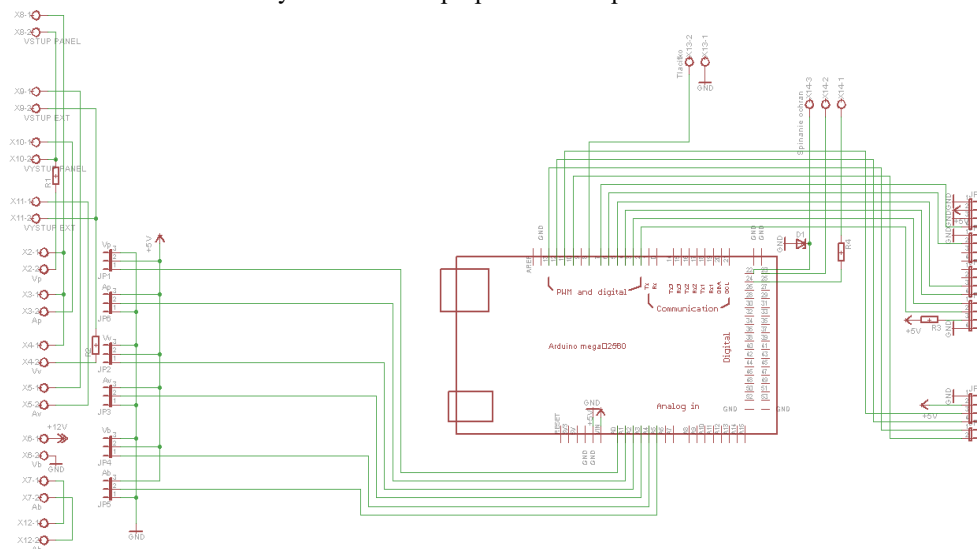


Obr. 3 Schéma zapojenia regulátora napájania pre Arduino

5) Dátovo informačné zariadenie

Keďže sa získavali hodnoty obvodu v reálnom čase a zároveň sa aj ukládali hodnoty na neskoršie porovnania, tak sa používal modul na SD kartu a LCD Displej 16×2, oba riadené cez Arduino. Na displeji sa zobrazovali hodnoty v reálnom čase a do SD karty sa ukládali hodnoty pre neskoršie využitie a vyhodnotenie.

Na displeji sa zobrazovali namerané hodnoty napätia, prúdu v reálnom čase na solárnom paneli a na výstupe pre napájanie zariadenia. Napätový senzor bol aj na batérii, kde ukazoval hodnoty vo voltoch a túto hodnotu vyhodnocoval a prepočítaval na percentá.

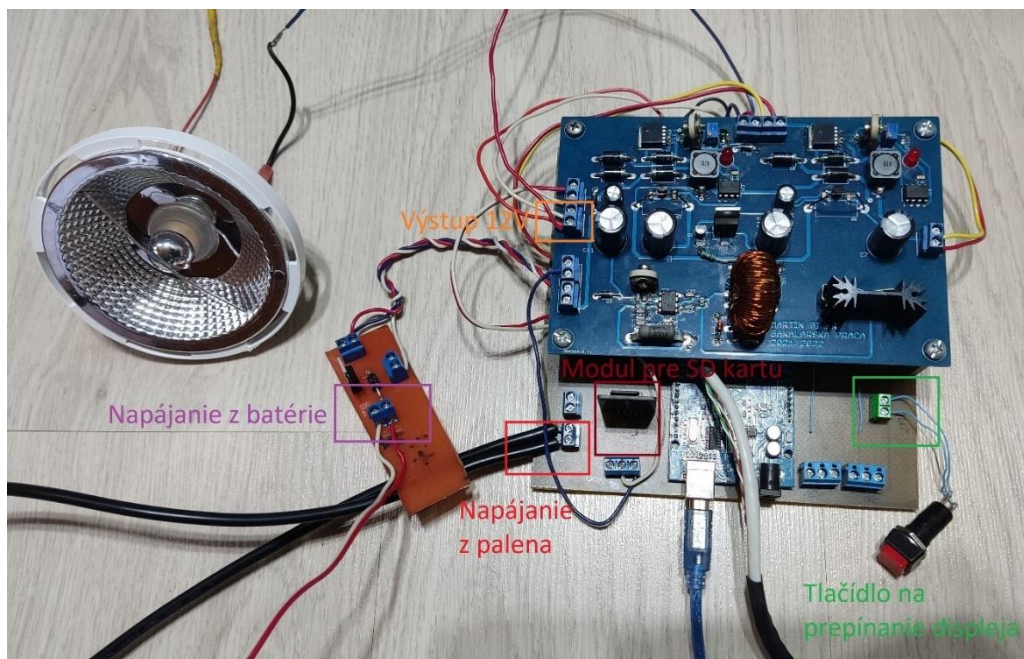


Obr. 4 Schéma zapojenia meracej a riadiacej dosky plošného spoja

B. Zostrojenie a vyhodnotenie zariadenia

Predtým, ako bolo pristúpené k osadeniu súčiastok, boli dané súčiastky podrobené meraniu. Ich výsledné hodnoty boli porovnané s hodnotami udávanými výrobcom. Tieto hodnoty sa líšili maximálne o 1,2 %. Nakoľko tieto hodnoty sa líšia minimálne, bolo možné pristúpiť k osadeniu týchto súčiastok.

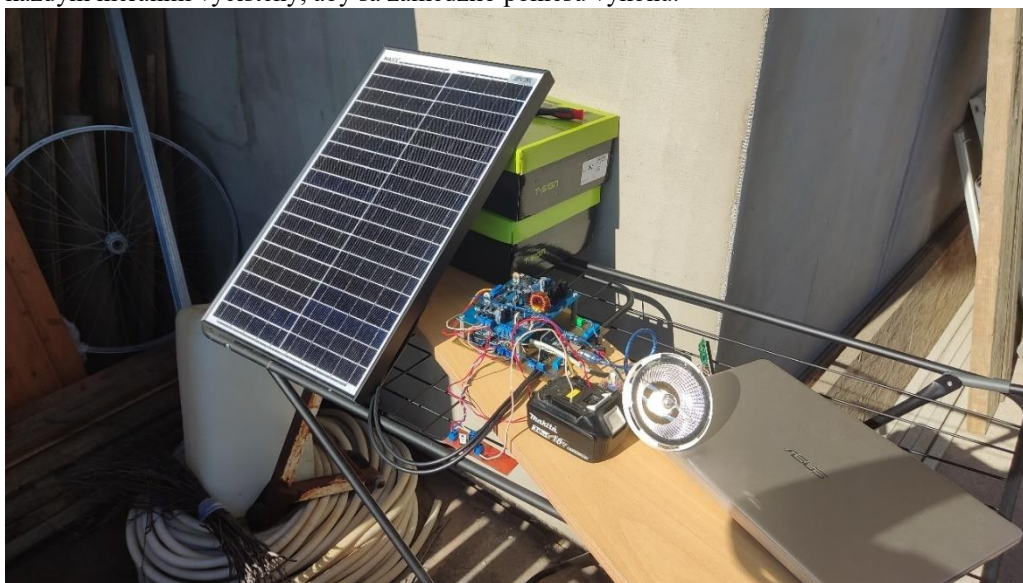
Všetky obvody boli navrhnuté v programe Eagle. Po nakreslení schém sa štandardným spôsobom postupovalo aj pri návrhu dosky plošných spojov. Vzhľadom na presnosť, jednoduchosť a potrebnú kvalitu, ktorá bola potrebná pri tejto doske, bola zvolená výroba vo firme na výrobu plošných spojov JLCPCB. Výrobné náklady a poštovné boli v porovnaní so slovenským výrobcom lacnejšie a kvalita výroby je porovnateľne rovnaká.



Obr. 5 Kompletné zhotovenie zariadenia

III. VÝSLEDKY MERANIA

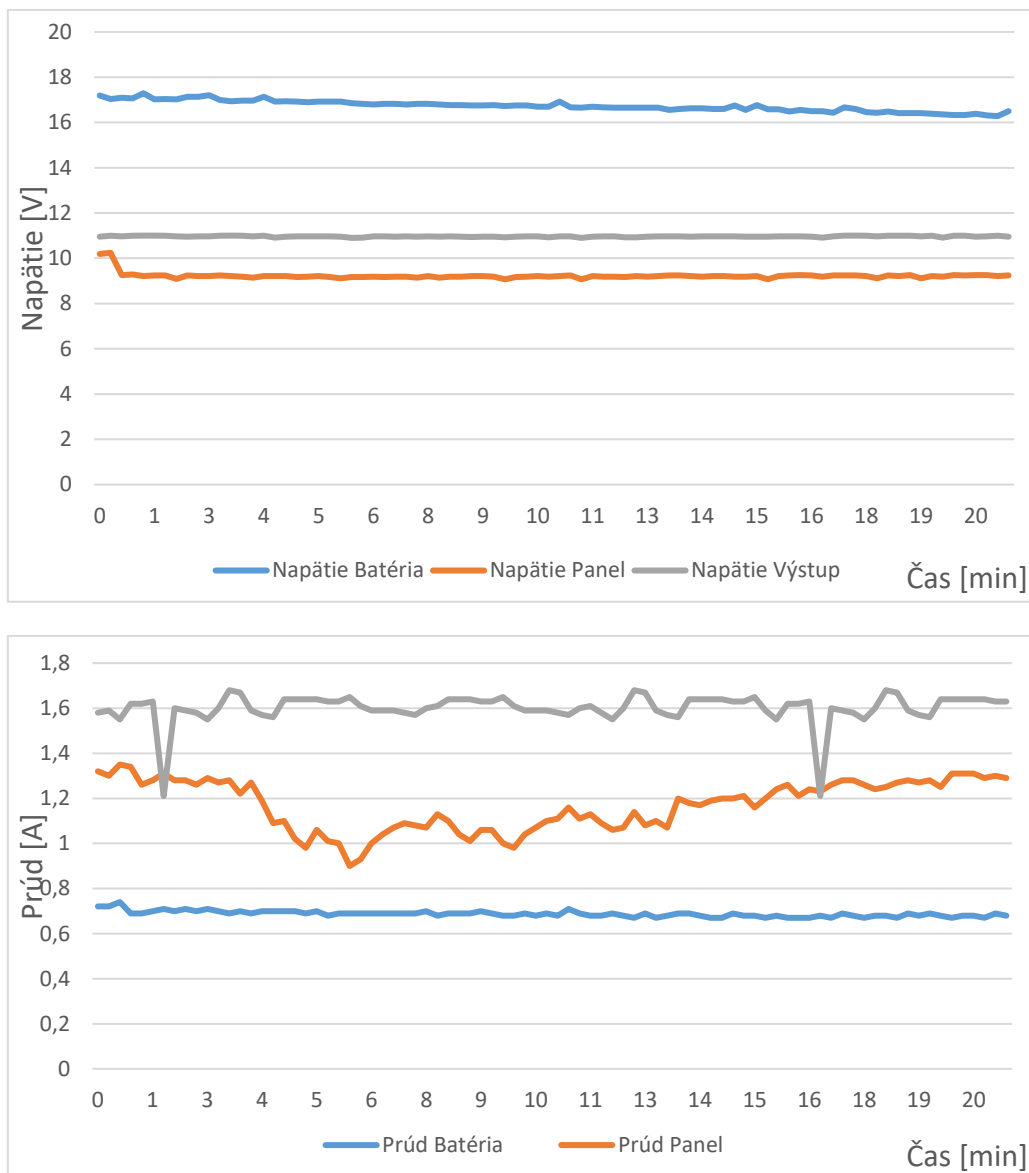
Všetky merania boli vykonávané na rovnakom mieste. Pre meranie bol vybraný taký čas a poloha, aby bolo možné získať čo najviac elektrickej energie z panela. Povrch panela bol každým meraním vyčistený, aby sa zamedzilo poklesu výkonu.



Obr. 6 Vykonávanie meraní

Pre zobrazenie výsledkov (pre účely tohto článku) bola ako záťaž použitá 4,5W LED žiarovka. Vďaka menšiemu odberu prúdu sa podarilo dosiahnuť jednak maximálny prúd a aj maximálne napätie, čiže maximálny výkon. Je potrebné poznamenať, že nabíjacia časť nebola využitá naplno. Do žiarovky tiekol maximálny prúd 0,38 A a z panel sa odoberal celkový prúd 0,6 A. Počas tohto merania bola vymenená batéria, preto aj výsledky obsahujú túto zmenu.

Vzhľadom na priaznivé slnečné podmienky bolo ešte otestované, ako bude panel reagovať, ak obidve žiarovky budú zapojené paralelne. Vzhľadom na fakt, že nie je možné dodať dostatočný výkon, tak sa nepodarilo rozsvietiť obidve žiarovky len pomocou panela. Práve preto bolo potrebné použiť akumulátor. Po zapojení akumulátora, na výstup sa dostáva stabilné napätie a panel vzhľadom na to môže dodávať maximálny prúd s danou hodnotou.



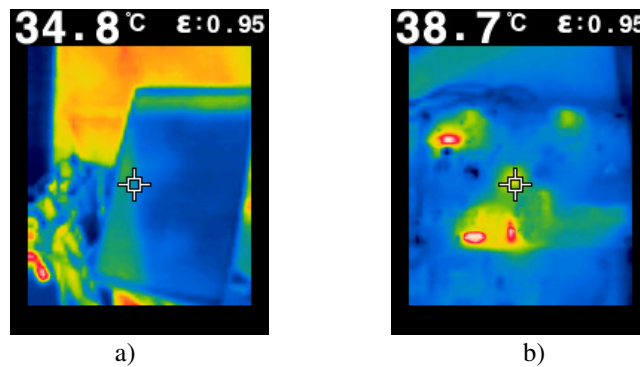
Obr. 7 Meranie na napájaní 4,5W + 15W žiarovky

Z uvedených meraní sa dá usúdiť, že celková účinnosť zariadenia klesá vtedy, ak sa nepoužíva batéria. Nakoľko fotovoltaický panel je prúdový zdroj, je veľmi závislý od pripojeného napätia. Práve z toho dôvodu je potrebný akumulátor, aby bolo možné udržiavať stabilné napätie pri vyšších a určitých odberoch. S akumulátorom bolo možné dosiahnuť výkon 20 W, avšak pri použití len čisto fotovoltaického panela sa na výstup podarilo dosiahnuť maximálne výkon do 10 W.

Celkové straty v obvode boli minimálne. Pri malých hodnotách výkonu boli celkové straty okolo 3 W, ktoré boli spôsobené napájaním Arduina. Pri vyšších výkonoch, napríklad pri meraní oboch žiaroviek, čiže výkonu 20 W, tieto straty činili okolo 5 W, kde sa straty prejavovali aj v obvodočných prvkoch. Tieto straty boli primárne tepelné. Vzhľadom na celkový výkon zariadenia, tieto straty sú pomerne veľké, nakoľko sa môžu vyšplhať do výšky 10 W. Vďaka termo-kamere bolo možné zaznamenať niektoré hodnoty teplôt pri meraní.

A. Meranie teplôt zostavy

Každý reálny elektrický obvod vytvára straty, ktoré sú najmä vo forme tepla. Elektronika konštruovaného zariadenia už počas meraní vykazovala nízke straty, a tým pádom aj nízke pracovné teploty. Všetky súčiastky a časti pracovali vo vyhovujúcej teplotnej hladine. Zároveň nebolo potrebné dodatočné chladenie na to, aby celkové zariadenie pracovalo spoľahlivo. V tomto prípade existuje veľká pravdepodobnosť, že ak by bolo použité aktívne chladenie v podobe ventilátorov, straty z napájania ventilátorov by mohli byť väčšie ako straty spôsobené teplom. Na meranie teplôt v obvode bol využitý princíp termovízie.



Obr. 8 a) Nameraná teplota na fotovoltickom paneli počas merania vyobrazená na termovíznej kamere; b) Namerané tepelné hodnoty na doske meničov počas napájania fotovoltickým panelom vyobrazené na termovíznej kamere

Pri bežnej prevádzke sa stredná teplota silovej dosky pohybovala v hodnotách teplôt okolitého prostredia. Iba samotné integrované meniče, ktoré regulovali napätie, vykazovali vyššie teploty. Na obrázku 8a v ľavom hornom rohu je zobrazený menič panela. Teploty boli merané v stave, kedy sa z panela nabíjali batérie. Teploty obidvoch regulátorov sa pohybovali v prevádzkových teplotách do 50°C. Tieto teploty sú zaznamenané pri priamom slnečnom žiarení.

IV. ZÁVER

Hlavným cieľom článku bolo navrhnúť a zostrojiť netypizovaný zdroj elektrickej energie, pre ktorý bol vybraný fotovoltaiický panel na princípe solárnej nabíjačky. Toto zariadenie dokáže dodávať elektrickú energiu s použitím batérie o výkone 20 W vzhľadom na vykonané merania.

Táto nabíjačka bola testovaná s rôznymi záťažami vo forme žiaroviek, nakoľko táto možnosť bola finančne a technicky nemenej náročná. Vzhľadom na to, že fotovoltický článok je monokryštalický a má veľké zníženie výkonu pri zhoršenom počasi, bolo potrebné vykonávať merania pri peknom a slnečnom počasi.

Vďaka výkonu, ktoré dokáže dodávať toto zariadenie a pomerne kompaktnému zostrojeniu, je toto zariadenie vhodné na to, aby sa vedelo zobrať do prírody, kde by napájal alebo nabíjal rôzne elektrické zariadenia. Keďže zariadenie využíva aj batériu, jeho celková účinnosť sa zvyšuje. Vďaka tomu pri nenapájaní žiadneho zariadenia, sa vie tento prebytok energie uchovať do batérie a následne využiť či už pri pomocnom napájaní s panelom alebo samostatnom napájaní alebo nabíjaní.

POĎAKOVANIE

Tento príspevok vznikol vďaka Agentúre na podporu výskumu a vývoja na základe zmluvy č. APVV-19-0576, ďalej Vedeckej grantovej agentúry Ministerstva školstva, vedy, výskumu a športu SR a SAV podporenej grantom VEGA 1/0757/21.

ZOZNAM POUŽITEJ LITERATÚRY

- [1] Mastný, P., „Obnoviteľné zdroje elektrickej energie“. [online], [cit.: 3.5.2022]. dostupné na: <https://www.mpo-efekt.cz/upload/7799f3fd595eeee1fa66875530f33e8a/cvut-2-oze.pdf>
- [2] Awwad, R., Das, R., Arabi, T., Hajj, H., „A fast charging Multi-C technique for mobile devices“, In: 2012 International Conference on Energy Aware Computing, 2012, pp. 1-7, doi: 10.1109/ICEAC.2012.6471022.
- [3] Nitta, N., „Li-ion battery materials: present and future“, In: MaterialsToday, Materials Today, Vol. 18, Issue 5, June 2015, pp. 252-264, doi: 10.1016/j.mattod.2014.10.040.
- [4] Neverman, A., „Everything You Need to Know About the 18650 Battery“, Common Sense Home, [online], [cit.: 4.5.2022]. dostupné na <https://commonsensehome.com/18650-battery/>.
- [5] Patočka, M., „Vybrané statě z výkonové elektroniky. Svazek II, Měniče bez impulsního transformátoru“, [online], [cit.: 1.5.2022]. dostupné na <https://vufind.mzk.cz/Record/MZK01-000752500>.
- [6] Binar, M., „Návrh netypizovaného energetického zdroja a jeho praktické využitie“. Bakalárska práca. Technická univerzita v Košiciach, Fakulta elektrotechniky a informatiky, 2022.

Ekonomické a environmentálne aspekty decentralizovanej výroby elektriny

¹Branislav VANGOR, ¹Samuel BUCKO

¹ Katedra elektroenergetiky, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

¹branislav.vangor@student.tuke.sk, ¹samuel.bucko@tuke.sk

Abstrakt — Elektrická energia tvorí základ pre ďalšie činnosti človeka. Elektráreň je jediným prvkom v elektrizačnej sústave, ktorý túto elektrickú energiu dokáže vyrobiť. Z toho dôvodu by prevádzka elektrárni mala byť čo najspoľahlivejšia. Obnoviteľné zdroje nám nedokážu poskytnúť spoľahlivú dodávku elektrickej energie, nakoľko ich výroba sa počas dňa mení. Dôležitým faktorom je inštalovať zdroje s čo najlepším plánovaním výroby elektrickej energie, aby bol dopad na elektrizačnú sústavu čo najmenší. Preto v predkladanom článku bola použitá jedna z metód multikriteriálneho rozhodovania na výber toho najvhodnejšieho variantu s čo najmenším vplyvom na celú elektrizačnú sústavu. Použitá metóda má názov analytický hierarchický proces.

KLúčové slová — decentralizovaná výroba elektrickej energie, obnoviteľné zdroje energie, multikriteriálne rozhodovanie, analytický hierarchický proces.

I. ÚVOD

Spočiatku bola výroba elektrickej energie založená na využívaní veľkých tepelných, jadrových či vodných elektrární. No postupom času, s cieľom znižovania znečistenia životného prostredia a vyššieho využitia obnoviteľných zdrojov elektrickej energie (OZE), najmä zo strany záväzkov voči EÚ, sa prechádza k tzv. decentralizovanej výrobe elektrickej energie. Vybudovanie veľkých výrobných kapacít využívajúcich OZE nie je najlepším riešením, a to hlavne z dôvodu kolísania ich výroby a následnej potreby veľkého regulačného výkonu. Decentralizovaná výroba – väčší počet výrobných kapacít rozmiestnených na určitom území, je lepším riešením, pretože zmena počasia nespôsobí taký výrazný pokles alebo zvýšenie výroby konkrétnych výrobných zdrojov. Neoddeliteľnou súčasťou procesu čo najlepšieho začleňovania OZE do sústavy sú, sa postupne stávajú kvôli možnosti obojsmerného toku elektrickej energie, inteligentné siete. V posledných rokoch stávajú často diskutovanou témou v energetike.

II. DECENTRALIZOVANÁ VÝROBA ELEKTRICKEJ ENERGIE

V minulosti bola elektrizačná sústava navrhovaná tak, aby zabezpečila prenos veľkého množstva elektrickej energie od elektrárne ku koncovým odberateľom. Ide o centralizovanú výrobu elektrickej energie, ktorá sa realizuje v malom počte no veľkých výrobných zdrojoch. Pri takejto výrobe elektrickej energie dochádza k znečisteniu životného prostredia nielen pri výrobe, ale aj pri ťažbe a doprave paliva do elektrárne. Umiestnenie týchto elektrární je často ďaleko od miesta spotreby, preto je po výrobe potrebné transformovať napätie na vyššiu napäťovú hladinu, aby sa znížili straty pri prenose elektrickej energie. V elektrických staniaciach blízko spotreby je znova potrebná transformácia na nižšiu napäťovú hladinu. Pôvodná myšlienka bola: čím sa postaví väčšia elektráreň, tým bude výroba efektívnejšia, ale vzhľadom na neustálu dekarbonizáciu energetiky sa energetické systémy začínajú postupne transformovať z centralizovaných na decentralizované. Ide hlavne o zvyšovanie využitia OZE, ktoré siete predstavujú menšie výkony v rozmedzí niekoľko stoviek wattov až do desiatok megawattov, ale počet týchto výrobných kapacít bude väčší v porovnaní s centralizovaným systémom. [1][2]

Pre decentralizovaný energetický systém, a inak nazývaný aj rozptýlená výroba či distribuovaná výroba, je charakteristické umiestnenie blízko miesta spotreby, kde je vyrobená elektrická energia dodávaná miestnym spotrebiteľom. Rozptýlená výroba pomáha k lepšiemu využitiu OZE, čo vedie k zníženiu spotreby fosílnych palív a zároveň k ochrane životného prostredia. OZE sa od konvenčných elektrární odlišujú nielen uhlíkovou neutralitou, ale aj ich modularitou, teda je ich možné postaviť ako samostatné jednotky alebo aj ako farmy s veľkosťou niekoľko gigawattov. Pri decentralizovanom systéme sú jednotlivé výrobné zdroje rozmiestnené

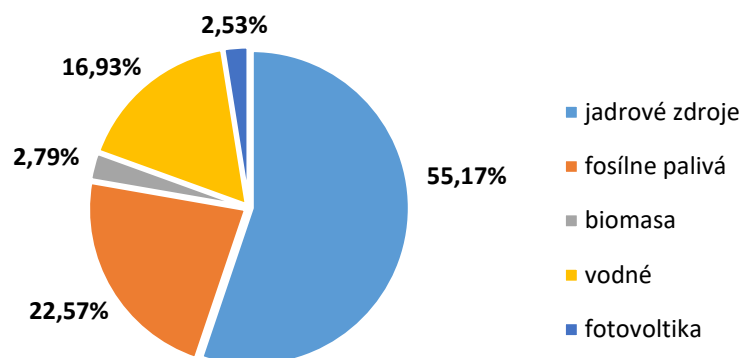
po celom území, bližšie ku koncovému užívateľovi, čím odpadá potreba prenosu elektrickej energie cez dlhé prenosové vedenia, pri ktorom dochádza k stratám výkonu. Výrobou elektrickej energie v menších množstvách v blízkosti spotreby dochádza k zvýšeniu účinnosti, menšiemu znečisteniu životného prostredia, zvýšeniu odolnosti siete a zníženiu nákladov prenosu na minimum. [3]

A. Výroba elektrickej energie z obnoviteľných zdrojov energie

Neustále zvyšovanie podielu výroby elektrickej energie z OZE je jednou z hlavných stratégií celej EÚ z dôvodu ochrany životného prostredia a boju proti zmene klímy. Krajiny EÚ dovážajú skoro 50 % primárnych zdrojov energie z krajín, ktoré nie sú členmi únie, aby pokryli svoju spotrebu. Práve technológie, ktoré využívajú OZE sú jedným riešením ako znížiť svoju závislosť od iných krajín, podporujú využívanie zdrojov, ktoré máme na Slovensku. Spočiatku boli OZE využívané najmä pre ostrovné prevádzky, no postupom času sa nerovnomernosť dodávky elektrickej energie dá vyriešiť použitím akumuláčnych systémov, ktoré síce zvyšujú investičné náklady, ale ich použitím je možné lepšie využiť elektrickú energiu, ktorá bola vyrobená z OZE. [4]

Problém s prepojitelnosťou OZE je najmä u veterných a fotovoltických elektrární, keďže základom pre fungovanie elektrizačnej sústavy je rovnováha výroby a spotreby elektrickej energie. A práve pri spomenutých OZE sa dokáže výroba elektrickej energie rapídne zmeniť v krátkom okamihu, pričom sústava musí byť schopná reagovať na tieto zmeny. Na vyrovnanie má prevádzkovateľ prenosovej sústavy k dispozícii podporné služby. Tieto podporné služby sú zdroje, ktoré dokážu meniť svoj výkon podľa požiadaviek dispečera. Nedostatočné množstvo regulačného výkonu sa môže odzrkadliť aj na rozvoji decentralizovanej výroby. Preto je potrebné udržať rovnováhu medzi výstavbou rozptýlených zdrojov a zdrojov, ktoré dokážu poskytovať podporné služby. Nekontrolovateľný rozvoj decentralizovaných zdrojov pre výrobu elektrickej energie môže spôsobiť aj nežiadúce zvyšovanie tokov elektrickej energie na cezhraničných vedeniach. [4]

Slovensko má domáce energetické zdroje hlavne v OZE a v jadrovej energetike a je takmer z 90 % závislé od dovozu primárnych zdrojov energie. Podiel, ktorým sa jednotlivé primárne energetické zdroje podieľali na výrobe elektrickej energie za rok 2020 je možné vidieť na Obr. 1. OZE sa spoločne podieľali na výrobe elektrickej energie 22,25 %. [5]



Obr. 1 Energetická skladba zdrojov SR za rok 2020 [5]

Členské štáty EÚ si môžu slobodne zvoliť legislatívne podmienky, ako docieľiť národné ciele v oblasti OZE. Cieľ EÚ je dosiahnuť podiel energie z OZE na hrubej konečnej spotrebe do roku 2030 aspoň 32 %. Slovensko si zvolilo cieľ 19,2 %. Podpora a rozširovanie využívania OZE je hlavnou témou pri návrhu trhu s elektrickou energiou. Dosiahnutie požadovaných cieľov EÚ do roku 2030 znamená, že až 50 % elektrickej energie budú vyrábať elektrárne využívajúce OZE. [5][4]

III. MULTIKRITERIÁLNE ROZHODOVANIE

Rozhodovanie ovplyvňuje následný priebeh ďalších udalostí, preto je potrebné venovať dostatočnú pozornosť rozhodovaciemu procesu. Pre rozhodovanie je dôležité mať dostatočné množstvo informácií o objektoch, ktoré vstupujú do rozhodovacieho procesu. Je dôležité sledovať hlavne tie vlastnosti objektov, ktoré sú dôležité pre rozhodovací proces. Multikriteriálne rozhodovanie je rozhodovacia analýza, ktorá vyhodnocuje viaceré možnosti riešenia „problému“ za pomoci definovaných kritérií. Pri hodnotení je možné sa stretnúť častokrát aj s protichodnými

kritériami. Multikritériálne rozhodovanie používa prakticky každý človek. Rozhodovací proces sa dotýka všetkých sfér, či už ide o energetiku, politiku, ekonomiku aj osobný život človeka.

Rozhodovacie problémy sú pri riešení popísané množinou kritérií, variantov a jednotlivých väzieb medzi kritériami a variantami. [6]

Pre vhodný výber metódy multikritériálneho hodnotenia je potrebné poznať: [6]

- o čom sa bude rozhodovať,
- aké ciele sa majú naplniť,
- hľadiská, z ktorých sa bude vychádzať,
- časové rozpätie počas ktorého bude výsledok multikritériálneho rozhodovania pôsobiť.

A. Analytický hierarchický proces

Jednou z metód multikritériálneho hodnotenia, ktorá určuje výsledné poradie variantov je metóda analytického hierarchického procesu (Analytic Hierarchy Process). Táto metóda sa označuje aj skratkou AHP či pomenovaním Saatyho metóda, pretože ju vymyslel práve Thomas L. Saaty na prelome 60-tych a 70-tych rokov 20. storočia. Je to najznámejšia a najviac využívaná metóda. Umožňuje rozdelenie zložitého problému na menšie celky, čím dochádza k lepšiemu priradeniu kritérií. Hlavný dôvod vysokého použitia tejto metódy spočíva v tom, že dokáže pracovať s kritériami, ktoré sú vyjadrené kvalitatívne alebo kvantitatívne. Metóda sa zaraďuje medzi metódy párového porovnávania, ale s rozdielom, že sa neurčuje iba preferencia jedného prvku nad druhým, ale určuje sa aj stupeň tejto preferencie. [7]

Rozhodovateľ má pre určovanie stupňa preferencie k dispozícii tabuľku so slovným popisom, podľa ktorého sa určuje významnosť porovnávaných prvkov. Základná stupnica párového porovnania – určenie preferencie pri metóde AHP, je v nasledujúcej Tab. 1. Využíva sa stupnica 1 až 9, pričom stupne 1, 3, 5, 7, 9 sú základné a majú aj slovné popisy. Je možné využiť aj medzistupne 2, 4, 6, 8, ktoré už nemajú slovné vysvetlenie a používajú sa vtedy, ak sa nie je možné rozhodnúť medzi hodnotami zo základnej stupnice. [7]

Tab. 1 Saatyho stupnica pre párové porovnávanie [6]

Stupeň hodnotenia	Definícia	Vysvetlenie
1	Rovnaká dôležitosť	Oba prvky sa rovnako podieľajú na intervencii cieľa.
3	Menšia dôležitosť jedného prvku	Skúsenosti a názory jemne preferujú jeden atribút pred druhým.
5	Podstatná alebo silná dôležitosť	Skúsenosti a názory silne preferujú jeden atribút pred druhým.
7	Demonštrovateľná dôležitosť	Jeden atribút je veľmi preferovaný a jeho dominancia je demonštrovaná v praxi.
9	Absolútna dôležitosť	Evidentné favorizovanie jedného atribútu pred druhým.
2,4,6,8	Stredné hodnoty medzi dvoma susednými posúdeniami	Ak je potrebný kompromis.

IV. APLIKÁCIA METÓDY ANALYTICKÉHO HIERARCHICKÉHO PROCESU NA VÝBER VHODNÉHO VARIANTU

Ako jednotlivé varianty na výrobu elektrickej energie sme považovali obnoviteľné zdroje energie:

- V1 : elektráreň využívajúca biomasu,
- V2 : veterná elektráreň,
- V3 : malá vodná elektráreň,
- V4 : fotovoltická elektráreň,
- V5 : geotermálna elektráreň.

Ďalším dôležitým krokom je určenie množiny kritérií. Kritériá pre hodnotenie jednotlivých variantov sú:

- K1 : investičné náklady na výstavbu elektrárne [€/MW],
- K2 : merné náklady na výrobu elektrickej energie [€/MWh],
- K3 : účinnosť výroby elektrickej energie [%],

- K4 : životnosť zariadenia [rok];
- K5 : možnosť presného plánovania výroby (predikcia),
- K6 : dostupnosť vhodného miesta pre výstavbu danej technológie,
- K7 : náročnosť výstavby,
- K8 : produkcia škodlivých látok [g/kWh],
- K9 : záber pôdy, prípadná montáž aj na už vybudované objekty,
- K10 : hluk,
- K11 : vizuálne zásahy do krajiny,
- K12 : negatívny vplyv na živočíchy.

A. Párové porovnanie

Výsledky párového porovnania sa zapisujú do Saatyho porovnávacej matice označovaná S , ktorá je zostavená v Tab. 2. Princíp tvorby matice je jednoduchý, porovnávajú sa prvky umiestnené v stĺpci a riadku:

- ak je prvok v riadku významnejší ako prvok v stĺpci, tak sa zapíše celočíselná hodnota významnosti ako sa určilo pri párovom porovnaní zo stupnice 1-9,
- v opačnom prípade ak je prvok nachádzajúci v stĺpci významnejší tak sa zapíše prevrátená hodnota stupňa významnosti, ktorý bol určený pri párovom porovnaní.

Matica je reciproká, teda obsahuje polovicu prvkov v prevrátenom tvare. Pri jej tvorbe je postačujúce vyplniť trojuholník nad alebo pod hlavnou diagonálou a opačná strana matice bude v prevrátenom tvare. Saatyho matica S_1 rozmeru 12x12 po zrealizovaní párového porovnania kritérií má tvar:

Tab. 2 Saatyho matica pre riešené rozhodovanie

	K1	K2	K3	K4	K5	K6	K7	K8	K9	K10	K11	K12
K1	1	7	1/3	1/5	1/9	1/7	1/4	1/8	1/5	1/2	1/6	1/8
K2	1/7	1	1/3	1/4	1/9	1/8	1/5	1/8	7	8	7	3
K3	3	3	1	7	1/8	1/7	8	1/9	2	7	5	8
K4	5	4	1/7	1	1/9	1/8	6	1/8	7	9	5	6
K5	9	9	8	9	1	1/2	8	6	7	9	8	9
K6	7	8	7	8	2	1	6	3	8	8	8	9
K7	4	5	1/8	1/6	1/8	1/6	1	1/8	9	3	5	8
K8	8	8	9	8	1/6	1/3	8	1	9	9	9	9
K9	5	1/7	1/2	1/7	1/7	1/8	1/9	1/9	1	7	8	8
K10	2	1/8	1/7	1/9	1/9	1/8	1/3	1/9	1/7	1	5	1/6
K11	6	1/7	1/5	1/5	1/8	1/8	1/5	1/9	1/8	1/5	1	1/8
K12	8	1/3	1/8	1/6	1/9	1/9	1/8	1/9	1/8	6	8	1

Ďalší postup bude podobný ako pri párovom porovnaní kritérií s rozdielom, že sa budú porovnávať jednotlivé varianty postupne z pohľadu zadaných kritérií. Kroky tvorby matice budú rovnaké, matica bude mať tvar 5x5 pretože je použitých 5 variantov. Spolu bude potrebné vytvoriť 12 matíc pre porovnanie variantov.

B. Výpočet váh

Po zrealizovaní všetkých párových porovnaní je možné pristúpiť k výpočtu jednotlivých váh, ktoré budú potrebné pri syntéze čiastkových hodnotení. Je možné použiť viacero metód pre výpočet váh. Východiskovou metódou pre výpočet váh je tzv. Saatyho postup, kde je potrebné vypočítať najväčšie vlastné číslo matice a vlastný vektor. Je to najzložitejší spôsob výpočtu a používa sa najmä v špecializovaných rozhodovacích programoch. Pre výpočet váh budeme používať Saatyho postup, kde na výpočet vlastného čísla a k nemu prislúchajúceho vlastného vektora využijeme program MATLAB, v ktorom sme zostrojili zdrojový kód pre výpočet váh kritérií a variantov.

Zo zadanej Saatyho matice pomocou príkazu „eig()“ v MATLAB-e bude identifikované najväčšie vlastné číslo matice a k nemu prislúchajúci vlastný vektor. Hodnoty vlastného vektora predstavujú nenormalizovanú váhu prvku. Vzájomné porovnanie váh kritérií je možné až po ich

normalizácii. Získané nenormalizované váhy kritérií bude teda potrebné normalizovať – zjednotiť dĺžku vektora, aby bol súčet jednotlivých váh rovný 1 a boli to nezáporné číselné hodnoty. Normalizáciu vektora dosiahneme vydelením každej nenormalizovanej váhy súčtom týchto nenormalizovaných váh.

C. Syntéza čiastkových hodnotení

Po vykonaní párových porovnávaní a výpočtu jednotlivých váh, je potrebná syntéza týchto jednotlivých hodnotení, aby sme získali výsledné poradie variantov. Postupne si do Tab. zapíšeme všetky normalizované váhy, ktoré sme získali výpočtom. Keďže každé kritérium má rôznu váhu (dôležitosť), celkové hodnotenie jednotlivých variantov získame ako vážený súčet váhy jednotlivého kritéria a váhy variantu z pohľadu určeného kritéria.

Tab. 3 Výsledné hodnotenie

Kritérium	Váha	Hodnotenie jednotlivých variantov				
		V1	V2	V3	V4	V5
K1 – investičné náklady na výstavbu	0,0230	0,0962	0,3420	0,0884	0,4756	0,0248
K2 – merné náklady na výrobu el. energie	0,0379	0,0877	0,2869	0,5141	0,0301	0,0812
K3 – účinnosť výroby elektrickej energie	0,0921	0,2346	0,0928	0,6186	0,0238	0,0301
K4 – životnosť zariadenia	0,0722	0,0427	0,0890	0,6544	0,0336	0,1803
K5 – plánovanie výroby	0,2356	0,3983	0,0413	0,3083	0,0304	0,2218
K6 – dostupnosť vhodného miesta na výstavbu	0,2078	0,4622	0,0849	0,0524	0,3649	0,0356
K7 – náročnosť výstavby	0,0580	0,0557	0,3323	0,1360	0,4555	0,0205
K8 – produkcia škodlivých látok	0,1799	0,4860	0,2738	0,1585	0,0566	0,0251
K9 – záber pôdy	0,0378	0,0813	0,3894	0,3524	0,0945	0,0824
K10 – hluk	0,0135	0,0808	0,0257	0,0536	0,5656	0,2743
K11 – vizuálne zásahy do krajiny	0,0149	0,0553	0,0243	0,6269	0,2253	0,0681
K12 – negatívne dopady na živočíchy	0,0273	0,0832	0,0480	0,0953	0,2869	0,4865
VÁŽENÝ SÚČET	1,0000	0,3174	0,1464	0,2717	0,1587	0,1059
PORADIE		1.	4.	2.	3.	5.

V. ZÁVER

Výpočtom sme dospeli k dvom výsledkom. Prvým z nich je, že pri rozhodovaní, z akého OZE bude realizovaná výroba elektrickej energie je potrebné prihliadať najmä na možnosť plánovania výroby, dostupnosti vhodného miesta na výstavbu a produkciu škodlivých látok. Ostatné z použitých kritérií majú vypočítanú nižšiu váhu a iba v malej miere ovplyvňujú výber vhodného variantu.

Druhým je výsledné poradie variantov. Na prvom mieste s najvyššou váhou sa umiestnila biomasa, a to hlavne z dôvodu možnosti plánovania výroby elektrickej energie, keďže pre výstavbu elektrárne na využívanie biomasy nie je potrebný lokálny prírodný potenciál, takže jej výstavbu je možné uskutočniť na hociktorom území. Nevýhodou však je dovoz biomasy, ktorý je potrebné následne obmedziť na čo najmenšiu oblasť. Na druhom mieste sa umiestnila malá vodná elektrárňa, ktorá síce potrebuje prírodný potenciál, ale je dobre prístupným a stálym zdrojom energie. Na treťom mieste sa nachádza fotovoltická elektrárňa, ktorá síce nepotrebuje špecifické prírodné podmienky a potenciál Slnka je dostupný, ale je to zdroj energie, ktorý je dosť ťažko predikovateľný. Na štvrtom mieste je veterná elektrárňa, ktorej výroba je taktiež ťažko predvídateľná a dokáže sa rýchlo zmeniť. Na poslednom mieste sa nachádza geotermálna

elektrárň, keďže na jej výstavbu sú potrebné vhodné prírodné podmienky. Pri tomto druhu OZE sú aj najvyššie investičné náklady, ktoré sú spojené najmä s vykonávaním vrto.

VI. POĎAKOVANIE

Tento príspevok bol podporený Agentúrou na podporu výskumu a vývoja na základe zmluvy č. APVV-19-0576, ako aj Vedeckou grantovou agentúrou Ministerstva školstva, vedy, výskumu a športu SR a SAV, podporený grantom VEGA 1/0757/21.

LITERATÚRA

- [1] KOLCUN, M. – VOJTEK, M. – POPOVEC, J.: Využitie akumulačných zariadení v elektrizačnej sústave [online]. In: Electrical Engineering and Informatics IX: proceedings of the Faculty of Electrical Engineering and Informatics of the Technical University of Košice. Košice: FEI TU, 2018. s. 664-668. [cit 2.2.2022] Dostupné na internete: <http://eei.fei.tuke.sk/data/EEI_IX.pdf>. ISBN 978-80-553-2713-6.
- [2] PÁL, D. – BEŇA, Ľ. – URBANŠKÝ, J.: Možnosti znižovania strát v sieti [online]. In: Electrical Engineering and Informatics X: proceedings of the Faculty of Electrical Engineering and Informatics of the Technical University of Košice. Košice: FEI TU, 2019. s. 226-231. [cit 2.2.2022] Dostupné na internete: <http://eei.fei.tuke.sk/data/EEI_X.pdf>. ISBN 978-80-553-3342-7.
- [3] UNESCAP: Decentralized energy system [online]. 2010. [cit 3.2.2022]. Dostupné na internete: <<https://www.unescap.org/sites/default/files/14.%20FS-Decentralized-energy-system.pdf>>.
- [4] JANÍČEK, F. a kol.: Obnoviteľné zdroje na Slovensku a ich vplyv na elektrizačnú sústavu. Bratislava: FEI STU, 2019. 221 s. ISBN 978-80-227-4966-4.
- [5] OKTE: Národný energetický mix [online]. 2022. [cit 9.2.2022]. Dostupné na internete: <<https://www.okte.sk/sk/zaruky-povodu/statistiky/narodny-energeticky-mix/>>.
- [6] OCELÍKOVÁ, E.: Multikriteriálne rozhodovanie. Košice : elfa, s.r.o, 2004. 87 s. ISBN 80-89066-28-3
- [7] JANDOVÁ, V.: AHP - její silné a slabé stránky : Diplomová práca. Olomouc : Univerzita Palackého v Olomouci – přírodovědecká fakulta, 2012. 114 s.

Riešenie bezpečnosti informačných technológií pomocou hier

¹Júlia DULINOVÁ ²Anton BALÁŽ, ³Emília PIETRIKOVÁ

^{1,2,3}Katedra počítačov a informatiky, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

¹julia.dulinova@student.tuke.sk ²anton.balaz@tuke.sk ³emilia.pietrikova@tuke.sk

Abstrakt – Práca sa venuje problematike gamifikácie a riešeniu bezpečnosti informačných a komunikačných technológií pomocou nej. Článok obsahuje analýzu súčasného stavu problematiky a príkladov gamifikácie v rôznych oblastiach života vrátane učenia bezpečnostných konceptov. Článok sa ďalej zaoberá návrhom vytvorenia praktického systému na podporu vzdelávania v oblasti bezpečnosti prostredníctvom gamifikácie. Návrh hry opisuje vypracovaná časť o jej schémach a následnej implementácii. Na záver článku sú spracované získané výsledky z dotazníka a získaných názorov na vytvorenú hru.

Kľúčové slová – bezpečnosť, dostupnosť, dôvernosť, gamifikácia, hra, integrita, utajenie

I. ÚVOD

Svet plný technológií dnes vyžaduje zmeny nášho správania a vykonávania aktivít. Zároveň so sebou prináša možnosť využitia týchto moderných mechanizmov v riešení vzniknutých problémov. V tomto kontexte sa naskytá príležitosť využitia herných prvkov k požadovanej zmene ľudského správania.

Hra ako taká je prirodzenou formou zábavy a zapojenie teórie hier predstavuje uľahčenie alebo zrýchlenie nudných procesov. Toto začlenenie je mimoriadne platným konceptom, čo sa týka učenia.

Odborníci na hry ich delia na aktivitu so štruktúrou a pravidlami, kde je potrebná snaha na dosiahnutie určitého cieľa (angl. game) a typ aktivity bez pravidiel, vnímaný ako spontánna improvizovaná aktivita bez cieľa (okrem cieľa zábavy z hry)(angl. play). Tieto aktivity vnímame ako jednotlivé celky hry alebo hrania sa. Naproti tomu sú elementy, ktoré tvoria tieto aktivity. Dané elementy charakterizujú gamifikáciu alebo hravý dizajn (podľa toho, do akého typu aktivity teoreticky spadajú).

II. GAMIFIKÁCIA A SÚČASNÝ STAV

Gamifikácia je charakterizovaná ako využívanie prvkov a mechanizmov herného dizajnu v nehermom prostredí[1]. Dôvodmi na riešenie teórie gamifikácie boli rastúce ovplyvňovanie spoločnosti počítačovými hrami a videohrami, ktoré vplývali na životné interakcie a aj to, že videohry sú určené viac pre zábavu ako užitočnosť[1]. Využívanie pojmu širšou verejnosťou začalo v ostatnom desaťročí, ale jeho prvý výskyt je zachytený v roku 2002 od Nicka Pellinga ako "zámerne škaredé slovo", ktoré sa vzťahovalo len na hardvér (hrou zrýchlený návrh používateľského rozhrania na rýchlejšie elektronické transakcie)[2]. V poslednej dekáde sa taktiež rozšíril počet článkov a publikácií o tejto téme[3], [4], čo značí nárast trendovosti a aktuálnosti témy.

Vo všeobecnosti môžeme povedať, že časť gamifikačnej teórie spočíva v rozložení hry ako hernej aktivity na jednotlivé elementy, ktoré sú neskôr pretavené do rozličnej štruktúry/obsahu v rôznych oblastiach. Za takúto oblasť môžeme považovať aj informačnú bezpečnosť, ktorej znalosti sú v súčasnosti potrebné u všetkých používateľov informačnej techniky.

Keďže hru hráme bez nejakej zjavnej pridanej motivácie, je možné vytvárať tento stav (keď sa venujeme nejakej činnosti s neporovnateľnou intenzitou a trvaním) aj umelo pri žiadaných poznatkoch a interakciách používateľov.

Gamifikácia je interdisciplinárnym prístupom, ktorý sa snaží motivovať používateľov k dosiahnutiu určitých behaviorálnych alebo psychologických výsledkov v rozličných oblastiach[4].

Riešenie gamifikácie je späté so zapojením používateľov do činnosti[5] a nadväzovaním na umelý konflikt definovaný pravidlami[6].

Aj keď sa to tak môže javiť, gamifikácia nie je len o premene bežných činností na hru. Ide o akési miešanie a kombinovanie pracovných procesov s hernými prvkami pre výslednú zvyšujúcu sa hladinu motivácie z plynúcej zábavy. Herné mechanizmy sú v gamifikovanom systéme preto, aby zabezpečili stránku zábavy z hry, ktorá je súčasťou systému[7].

Gamifikácia a hra nie sú synonymá, pretože hra je primárne určená na zábavu. Avšak tá je neoddeliteľnou súčasťou gamifikovaného systému a týmto posúva daný proces dopredu rýchlejšie.

Príbuznými pojmami ku gamifikácii sú vážne alebo seriózne hry (serious games) a učenie založené na hre (gamebased learning). Niektorí rozlišujú gamifikáciu z dôvodu rozsahu použitia hracích prvkov (kým seriózna hra je kompletná hra v presne definovanom hernom priestore na nezábavne účely, gamifikovaná aplikácia používa len prvky herného dizajnu mimo definovaný herný priestor). Iní zase seriózne hry zaradzujú ako podmnožinu ku gamifikácii[8].

Typy gamifikácie

Gamifikácia môže byť štruktúrna alebo obsahová. Pri štruktúrnej gamifikácii je kladený dôraz na prechádzanie obsahom a odmenenie učenia (použitie rebríčkov (leaderboards), sledovanie učebného progresu, zdieľanie svojich úspechov a pod.). V rámci použitia tohoto typu je možné pridať aj elementy príbehu a postáv bez zmeny obsahu.

Obsahová používa herné prvky s úmyslom meniť obsah, aby sa viac podobal hre (príbehové elementy na splnenie úlohy, výzvy namiesto nezaujímavých cieľov a pod.). Zaraďujeme tu pridávanie príbehových elementov na splnenie úlohy alebo výziev namiesto zoznamu cieľov.

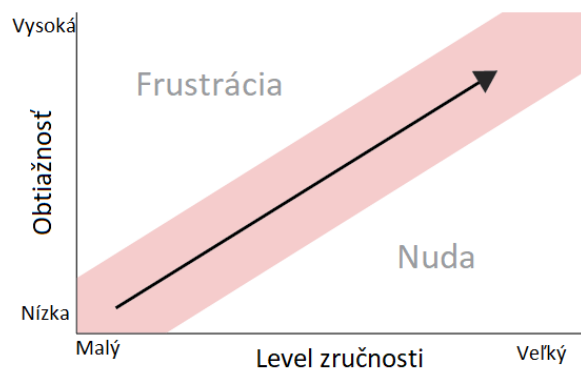
Je potrebné poznamenať, že odmeny (rewards), body (points), rebríčky (leaderboards), dosiahnuté úspechy (achievements) a levely nie sú samé o sebe gamifikáciou[9].

Dôvody použitia

Nesprávne dôvody pre použitie gamifikácie sú častokrát povrchné. Radí sa tu vyriešenie každého problému vo všetkých smeroch, nenáročné učenie s ľahkou implementáciou alebo aktuálne trendy. Gamifikácia má pozitívny efekt, ak je vyžadovaná interaktivita vedúca k učeniu, vysoké poznanie zapájaného kolektívu, vhodne zvolená forma gamifikácie, nutnosť autentickej praxe (simulačné prostredia), prekonanie neagnažovanosti alebo vypojenie z činnosti, kolektívu, témy a iné.

Môžeme zachytiť zjavnú podobnosť hier, vernostných programov a gamifikácie. Majú spoločné dobrovoľné zapojenie, využitie niektorých herných štruktúr a podobne[2]. Rozdiely panujú v stupni záujmu a cieľoch. Hry sú jednoduchou zábavou, vernostné odmenové programy sú transakciou a gamifikácia je umocňovaním motivácie prostredníctvom emočného levelu. Čo sa týka cieľov, pri gamifikácii je zjavné zdieľanie spoločného cieľa. Podobne ako u vážnych hier, ani tu nie je teória vždy jednotná, preto záleží na cieľnom záujme a účele implementácie.

Dôležitým aspektom je motivácia ako proces dodávajúci energiu, smer alebo zmysel konaniu a správaniu. Jej najvyššia hodnota sa dosahuje, ak aktivita nie je príliš zložitá, ale ani ľahká. Spadá do nej v rámci empirickej skúsenosti koncentrácia a potešenie súčasne (zanepřázdenie činnosťou a strata pojmu o čase). Pri definícii troch tokov (tok nudy, prúdový tok (flow) a tok frustrácie) z výskumu Mihaly Csikszentmihalyiho je cieľom, aby aktivita spadala medzi kanály nudy a frustrácie do prúdového toku[10].



Obr. 1 Teória toku a jej rozloženie

Elementy hier

Pri rozdeľovaní hry sa prvky prelínajú a ich detaily sa líšia od autora k autorovi. Podľa Jesseho Schella je základom elementárny tetrad ako spojenie mechaniky, príbehu, estetiky a technológie[11]. Karl Kapp uvádza vo všeobecnosti spomína mechaniku, estetiku, získanie záujmu ľudí, motiváciu k aktivite, podporu učenia a riešenie problémov[12]. Detailnejšie uvádza medzi použiteľnými gamifikačnými elementami abstrakciu reality, ciele, pravidlá, konflikt (súťaženie, spolupráca), čas, odmenové štruktúry, spätnú väzbu, levely, rozprávanie príbehu, krivka záujmu, estetika a opätovné hranie hry[8]. Uspokojivú gamifikáciu vytvorí vzájomný vzťah medzi týmito prvkami (použitie jedného alebo dvoch prvkov nám nezaručí hru s pohlcujúcim prostredím a použitie všetkých automaticky nevytvorí dokonalú hru).

Existujúce riešenia

Iniciatíva Zábavná teória od Volkswagenu (The Fun Theory) pred desaťročím motivovala ľudí k akcii prostredníctvom rôznych projektov ako "klavírne schody", recyklačná nádoba ako arkádna hra, odpadkový kôš so zvukom a i.

Dnešné aplikácie na zlepšenie životného štýlu ako Fitbit, Nike Run Club alebo Zombies, Run! v spojení zbierania dát, následnej analýzy, spracovania a zobrazenia používateľovi vyústi v ďalšie používanie[13], [14], [15]. Najpopulárnejšia platforma na výučbu jazykov Duolingo je úspešná aj vďaka dobre analyzovanej a spracovanej forme gamifikácie[16]. Hry s potenciálom naučiť (napr. Kerbal Space Program (fyzika) alebo hra Variants:Limits (matematika)[17], [18]) primárne nie na to určené, ale ich spracovanie ich k tomu predučuje.

Využitie gamifikácie neobišlo ani reklamy a marketing. Jej úspešne použitie bolo zaznamenané napr. spoločnosťami M&M's, CocaCola, Wendy's a iné[19], [20].

Elearningové platformy sú špeciálnym odvetvím učenia založeným na gamifikácii ako napr. Kahoot, Archy Learning alebo Quizizz.

V rámci gamifikácie bezpečnosti sú existujúcimi riešeniami napr. Project Ares[21](školenia a hodnotenie kybernetickej bezpečnosti), Living Security (riešenia pre riadenie bezpečnostných rizík ľudským manažmentom)[22], NOVA labs (podporujúce vedecké skúmanie)[23] a ďalšie.

III. NÁVRH A IMPLEMENTÁCIA RIEŠENIA

Riešenie je navrhnuté ako hra zahrňujúca v sebe prvky hry a zároveň učenia bezpečnostnej teórie. Tieto dve pomyselné roviny spojíme v jednoduchú "skákačku". Jej mechanika a estetika bola riešená pomocou statickej webovej stránky s použitím HTML, CSS a vanilla Javascriptu.

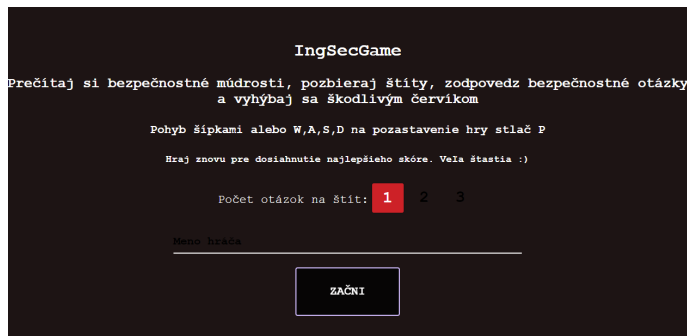
Hra mala spĺňať úlohy učenia, testovania, odmeňovania hráča bodovo a priebežne mu zobrazovať jeho stav počas celej hry (myšlienka "juicy" systému). Tomuto progresu bude priradený vyčísliteľný výsledok. Hra využije levelové rozloženie s rozdelením cieľov do menších skupín, ktoré sa dosiahnu v postupnej štruktúre. Nové poznatky budú hráčovi poskytované v textovej forme pred začiatkom levelu. Na základe nich bude možné zvoliť správnu odpoveď.

Dôležitým konceptom je stav hry. Na základe toho je možné prepájanie hernej a náučnej časti. Herne je najdôležitejší objekt hráča, ktorého pohyb zodpovedá vstupu s klávesnice. Spätná väzba je zobrazením jeho konfliktov s inými objektmi, ktoré ovplyvňujú získané skóre inkrementovaním alebo dekrementovaním (štit alebo červ). Pohyb je rovnako obmedzený stenami. Funkcia kolízie dvoch objektov je vykonávaná neustále v hracej slučke v hlavnej funkcii hry. V rámci levelu je určenou náplňou hry je zozberanie všetkých štítov. Tieto odmeny nám inkrementujú skóre a zároveň presmerujú hru na učebnú časť. Otázky sú navrhnuté unikátne v rámci levelu. Obsahovo sú tvorené z časti predmetu Bezpečnosť informačných a komunikačných systémov (študijného programu Informatika na FEI TUKE), ide teda o inžiniersky level poznatkov (odtiaľ aj názov IngSecGame (inžinierska bezpečnostná hra)). Odpovede môžu a nemusia byť zistené na začiatku levelu v náučnom texte. Ich zobrazenie je náhodné v rámci levelovej tematiky. V prípade nezískania dostatočnej informácie by tento stav mal navádzať na opätovné hranie kvôli ďalším vedomostiam.

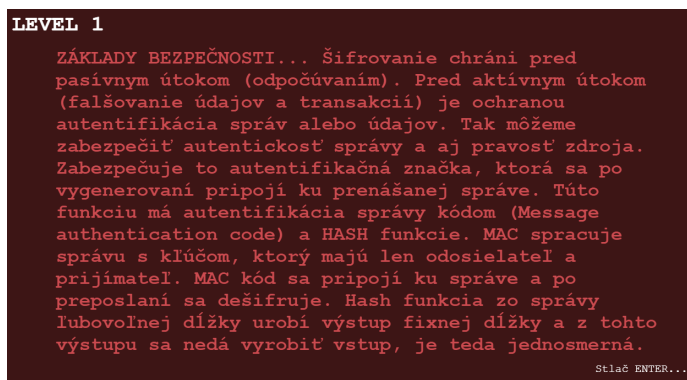
Hra začína úvodnou obrazovkou, ktorá vysvetľuje pravidlá a pokyny hry. Po ich prečítaní si hráč vyplní svoje meno, pod ktorým chce hrať. Môže si tu vybrať koľko otázok chce dostávať na jednu zozbieranú odmenu, čo znamená, že pri vyššom čísle je možný zisk viacerých bodov v hre.

Po spustení hry je zobrazený učebný text, ktorý sa zobrazí podľa terajšieho levelu. Po jeho prečítaní je hráč vpustený do levelu, kde bude mať za úlohu zozbierať odmeny v podobe štítov. Zároveň sa snaží vyhýbať nepriateľským červom, ktoré mu pri kontakte budú skóre znižovať.

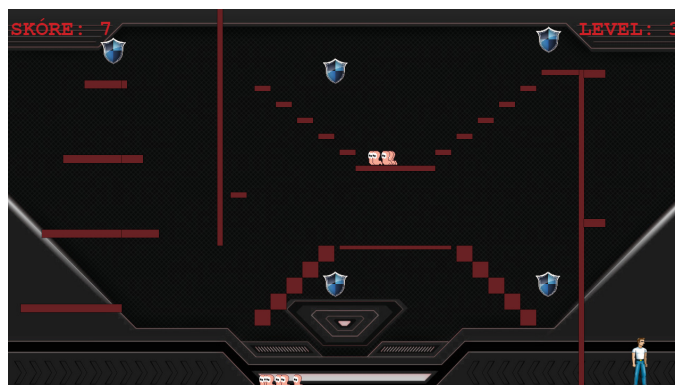
Ak sa hráčovi podarí zobrať štít, je mu položených toľko otázok, koľko si vybral na začiatku hry v úvodnej obrazovke. Otázka pozostáva zo znenia a 4 možných odpovedí, z ktorých je jedna správna. Po označení odpovede je farebne naznačená jej správnosť/nesprávnosť. Následne je implementovaná možnosť nápovedy (v pravom hornom rohu otázky), ktorá má za úlohu nasmerovať hráča ku správnej odpovedi alebo poskytnúť informáciu, ktorá logicky okliešti výber



Obr. 2 Úvodná obrazovka hry



Obr. 3 Príklad náučného textu na začiatku levelu

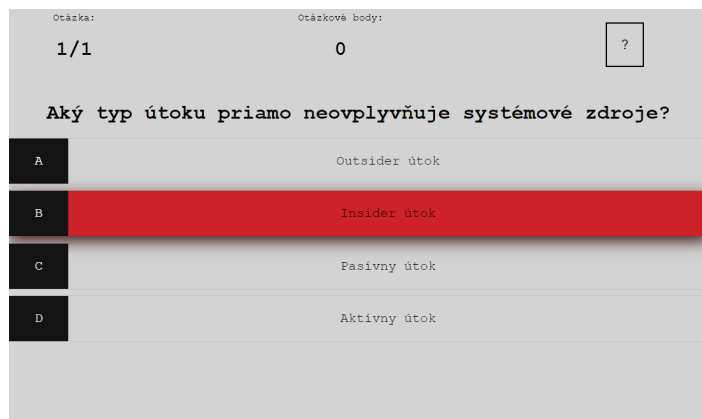


Obr. 4 Zobrazenie tretieho levelu v jeho úvode

z daných možností. Aby sme sa v hre vyhli frustračnému toku, za nesprávne odpovede sa body neodčítavajú. Rovnako použitie nápovedy nie je v žiadnej forme penalizované. Po zodpovedaní všetkých otázok sa hráč vráti k zobrazeniu hráča, ktorý môže pokračovať v love štítov. Ak všetky štíty boli zozbierané, zobrazí sa obrazovka oznamujúca koniec levelu a bodový stav po ňom. Pokračuje sa na ďalší level so zobrazením učebného textu z inej bezpečnostnej problematiky. Ak bol daný level posledný, hráč obdrží textové hodnotenie podľa počtu správne zodpovedaných otázok a celkové skóre. Z tejto koncovkej obrazovky hry je možné prejsť na vykreslenie rebríčka najlepších. Hráč musí obdržať kladné skóre aby mohol byť do rebríčka zaradený. Okrem skóre je hráč určený menom, ktoré zadal v úvode, dátumom a časom svojej hry. Následne je možné hru reštartovať. V takom prípade ide o presmerovanie na úvodnú obrazovku bez stôp predchádzajúcej hry (s výnimkou zapísaného skóre v lokálnom úložisku prehliadača).

Implementačné zhrnutie

Mechanika hry bola implementovaná v spojení s teoretickou časťou. Hráčovi je oznámené ovládanie a pravidlá hry, čo vytvára hranice v ktorých pôsobí na dosiahnutie jej cieľa. Z detailnejšieho hľadiska bolo v hre integrované použitie bodov, postup v leveloch, rebríčok



Obr. 5 Príklad kvízovej otázky s nesprávnou odpoveďou

Poradie	Hráč	Skóre	Dátum
1	Pasion de Gavilanes	87	30. 3. 2022, 19:16:42
2	Ronacio	83	29. 3. 2022, 14:01:11
3	Anonym	38	29. 3. 2022, 15:02:37
4	Misia	29	22. 3. 2022, 15:32:26
5	Anonym	26	24. 3. 2022, 10:23:41
6	Anonym	26	28. 3. 2022, 19:25:48
7	Anonym	25	24. 3. 2022, 10:13:39
8	Anonym	25	24. 3. 2022, 10:18:33
9	Anonym	25	24. 3. 2022, 10:28:47
10	Montes	22	29. 3. 2022, 14:13:31

Obr. 6 Zobrazenie rebríčka hráčov s najväčším počtom bodov na konci hry

najlepších lokálnych skóre. Tieto prvky sú naviazané na využitie spätnej väzby (zobrazenie aktuálneho bodového stavu a levelu). Neskôr dodaná nápoveda mala za úlohu udržať hráča mimo toku frustrácie v kanáli záujmu. Za prvok estetiky sme v hre využili pozadie, ktorého farba sa menila v závislosti od levelu a hlavne zobrazenie hráča (aj objektov štítu a červa). S spojením s mechanikou pohybu estetiku zabezpečuje aj ich animácia.

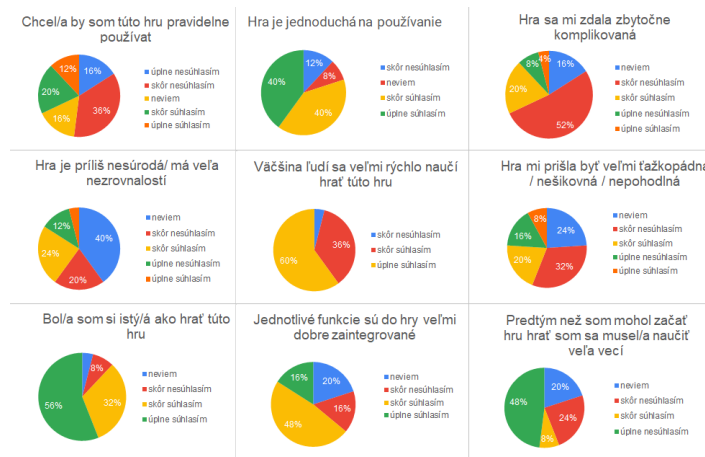
Týmto riešením je možné učiť každú teoretickú znalosť, stačí len vytvorenie učebného textu a otázok.

IV. VYHODNOTENIE

Vyhodnotenie riešenia bolo vykonané dotazníkovou spätnou väzbou. Otázky boli položené študentským respondentom o použiteľnosti a použitých prvkoch hry. Použiteľnosť sme testovali dotazníkovými otázkami miery použiteľnosti (SUS). V priemere boli respondenti neutrálni pri pravidelnom využívaní alebo komplikovanosti hry. Skôr súhlasili s jednoduchosťou použitia a dobrým zaintegrovaním funkcií v hre. Nerozhodne sa vyjadrovali aj o nezrovnalostiach v hre, jej ťažkopádnosti a nepohodlnosti. Panoval úplný súhlas, že sa väčšina ľudí naučí hrať túto hru rýchlo. Takisto skôr súhlasili s istotou, ktorú mali pri jej hraní. Pri hraní hry skôr nesúhlasili s tým, že sa museli naučiť mnohé veci, aby ju mohli hrať. Podľa priemerného hodnotenia každej odpovede sme po tradičnom výpočte dosiahli mieru použiteľnosti systému 68,2. Pri dolnej hranici použiteľnosti 68 sme teda dosiahli hraničný výsledok. Tvrdenie ohľadom ohraničenia hry časom odhalilo, že polovica by pridanie časomieru vnímala pozitívne a druhej by to prinieslo neželaný stres. Toto polovičné rozdelenie, len potvrdilo odhad dôležitosti individuálnych preferencií hráčov.

V zhrnutí ďalších zistení bola dĺžka hry primeraná, akurátna obtiažnosť hry pre väčšinu (36%), dostatočné vysvetlenie hry (92%), malý vplyv vstupu hráča (40%) a nevšímnutie si prvku nápovedy (44%). Hodnotenie hry bolo v 68% ako Dobré, celková spokojnosť s výsledkom bola 76%. Podľa väčšiny by hru zlepšilo pridanie herných prvkov ako odznaky, iné odmeny, pridané body za viac správnych otázok v rade a i.

Gamifikácia takouto formou by pomohla výučbe bezpečnosti podľa 80% respondentov. Medzi študentmi existuje dopyt po takomto učení (dokazuje to aj 60% podpora kombinovanej (tradičnej a gamifikovanej) prípravy, ak by bola možná).



Obr. 7 Diagramové zobrazenie otázok pre mieru použiteľnosti

V. ZÁVER

Hlavný cieľom článku bolo zoznámiť čitateľa s problematikou gamifikácie a vytvoreným praktickým riešením. Teória hier a jej elementov je v istých smeroch mnohznačná, čo sťažuje pochopenie a následné implementovanie tejto zdanlivo jednoduchšej témy. Jej využívanie je v pracovných, školských aj iných náučných sférach na vzostupe. Spôsoby a formy sú bohaté, ak ide o správne využitie a dosiahnutie cieľov. Bližšie informácie o gamifikácii, teórii hier, existujúcich riešeniach je možné nájsť v uvedenej použitej literatúre.

POĎAKOVANIE

Článok bol podporený v rámci projektu 002TUKE4/2021 Implementácia moderných metód a foriem výučby v oblasti kyberbezpečnosti k požiadavkám praxe.

LITERATÚRA

- [1] S. Deterding, R. Khaled, L. Nacke, and D. Dixon, "Gamification: Toward a definition," 01 2011, pp. 1215.
- [2] B. Burke, "Gamify: How gamification motivates people to do extraordinary things," 2014.
- [3] J. Hamari, J. Koivisto, and H. Sarsa, "Does gamification work? a literature review of empirical studies on gamification," in *2014 47th Hawaii International Conference on System Sciences*, 2014, pp. 30253034.
- [4] S. Stieglitz, C. Lattemann, S. RobraBissantz, R. Zarnekow, and T. Brockmann, *Gamification Using Game Elements in Serious Contexts*. Gewerbestrasse 11, 6330 Cham, Switzerland: Springer International Publishing, 2017.
- [5] F. Groh, "Gamification: State of the art definition and utilization," *Proceedings of the 4th Seminar on Research Trends in Media Informatics*, pp. 3946, 01 2012.
- [6] K. Salen and E. Zimmerman, *Rules of Play Game Design Fundamentals*. Massachusetts London, England: The MIT Press Cambridge, 2004.
- [7] L. Wood and T. Reiners, *Gamification*, 01 2015, pp. 30393047.
- [8] K. M. Kapp, *The Gamification of Learning and Instruction: GameBased Methods and Strategies for Training and Education*, 1st ed. Pfeiffer & Company, 2012.
- [9] K. Kapp, "The gamification of learning and instruction fieldbook: Ideas into practice." John Wiley & Sons, 2013.
- [10] R. Buchanan and M. Csikszentmihalyi, "Flow: The psychology of optimal experience," *Design Issues*, vol. 8, no. 1, p. 80, 1991.
- [11] J. Schell, *The Art of Game Design: A Book of Lenses*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2008.
- [12] K. Bell, *Game On!: Gamification, Gameful Design, and the Rise of the Gamer Educator*. Johns Hopkins University Press, 2017.
- [13] N. Turner, "How fitbit have nailed gamification & 10 lessons you can learn from them," 06 2018. [Online]. Available: <http://www.uxforthemasses.com/10gamificationlessons/>
- [14] J. Vanhaesebroeck, "Gamification examples — nike run club," 08 2021. [Online]. Available: <https://strivecloud.io/blog/appgamification/gamificationexamplesnikerunclub/>
- [15] "Zombies, run!" [Online]. Available: <https://zombiesrungame.com/>
- [16] J. Bilham, "Case study: How duolingo utilises gamification to increase user interest," 07 2021. [Online]. Available: <https://raw.studio/blog/howduolingoutilisesgamification/>
- [17] S. Oberdörfer and M. Latoschik, "Effective orbital mechanics knowledge training using game mechanics," 09 2018.
- [18] "Variant: Limits." [Online]. Available: <https://triseum.com/variantlimits/>
- [19] "Gamification of product advertising: M&ms," 07 2019. [Online]. Available: <https://corp.gamitize.com/2017/01/22/gamificationofproductadvertisingmms/>
- [20] "Wendy's marketing gamification strategies gamify clients." [Online]. Available: <https://www.gamify.com/wendys>
- [21] "Project ares, the gamified learning platform that makes teaching cybersecurity within reach." [Online]. Available: <https://projectares.academy/>
- [22] "Solutions." [Online]. Available: <https://www.livingsecurity.com/cybersecurityhumanriskmanagement>
- [23] "Cybersecurity — nova labs — pbs." [Online]. Available: <https://www.pbs.org/wgbh/nova/labs/lab/cyber/>

Analýza regresných metód z hľadiska vhodnosti na detekciu trolov

¹Kristína MACHOVÁ, ²Martin KAŇUCH

^{1,2} Katedra kybernetiky a umelej inteligencie, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach

¹kristina.machova@tuke.sk, ²martin.kanuch@student.tuke.sk

Abstrakt — Práca sa venuje problému rozpoznávania internetových trolov na sociálnych sieťach a poukázať na možné dôsledky, ktoré títo používatelia spôsobujú. Práca ponúka teoretický prehľad rôznych typov trolov ako aj problematiky ich rozpoznávania pomocou regresných metód. Článok analyzuje rôzne regresné metódy z hľadiska použitia na riešenie pomenovaného problému. Práca popisuje budovanie modelov učením na dátových množinách extrahovaných zo sociálnej siete po predspracovaní. Napokon budú jednotlivé modely otestované a analyzované z hľadiska vhodnosti na budovanie detekčných modelov na detekciu trolov.

Kľúčové slová — troling, detekcia trolov, regresné metódy, strojové učenie, dolovanie textu

I. ÚVOD

Prístup ľudí na internet sa v posledných rokoch stal nedielnou súčasťou nášho života a ľudia sú schopní byť pripojení k internetu takmer neustále. Vznik sociálnych sietí spôsobil to, že ľudia majú viac dôvodov ostávať pripojení dlhšie a dlhšie a sledovať čo sa deje v živote ich priateľov, čo sa práve deje vo svete a rýchlo sa informovať o potrebných faktoch. V západných krajinách a krajinách ďalekého východu sa zdá ťažšie byť odpojený ako byť pripojený k internetu. Už od začiatkov internetových fór a četovacích miestností sa rýsoval problém dôveryhodnosti informácií na internete. Častejším sledovaním sociálnych sietí a internetových fór ale zvyšujeme riziko toho, že narazíme na nepravdivé informácie a ľudí, ktorí sa nazývajú internetoví trolovia, ktorí sa vyžívajú v tom, že spôsobujú nezhody medzi používateľmi. Byť pod palbou internetového trola môže byť veľmi nepríjemným zážitkom, ktorý niektorým ľuďom, hlavne mladistvým, môže spôsobiť až vážne psychické problémy. Tento článok sa zameriava na odhalenie, kto sú títo ľudia, čo robia, ako ich detegovať a ako im odolávať.

II. TROLING A TROLOVIA

Troling je typický vytváraním sporov na internete, vyvolávaním hádok alebo rozruchu medzi ľuďmi, zverejňovaním „zápalových správ“ alebo správ mimo témy v online komunite. Trol na sociálnych sieťach je v zásade niekto, kto zámerne hovorí niečo kontroverzné, aby dostal reakciu z ostatných používateľov [1]. Komentár zverejnený trolom obsahuje takmer vždy negatívny obsah. Ale nie každý negatívny komentár, ktorý sa na prvý pohľad javí ako trol, ním naozaj je. Treba vedieť rozlíšiť medzi správou ktorá je negatívna z dôvodu reálnej frustrácie používateľa alebo pokusu používateľa o nejaký vtíp, satiru a sarkazmus a vedieť rozlíšiť ktoré komentáre sa naozaj snažia vyvolať hádku a zahltiť cieľového používateľa rovnakými správami a spamom.

Internetových trolov môžeme nájsť takmer úplne všade, na každej sociálnej sieti, fóre alebo v komentároch a recenziách online obchodov, v emailoch a blogoch. Šanca že ihneď narazíme na nejakého trola je relatívne nízka ale na fórach ako je napríklad Reddit, 4Chan, 8Chan je šanca oveľa vyššia. Dôvodom je to, že tieto fóra poskytujú anonymitu používateľom a potom títo trolovia nie sú nijako obmedzovaní v trolingu. Problém nastáva keď trolovia začnú rozširovať hoaxy a nepravdivé konšpiračné teórie, ktoré sa nedajú ihneď vyvrátiť, pretože neexistujú priame dôkazy na ich vyvrátenie alebo dôkazy existujú, no bežní používatelia k týmto informáciám nemajú prístup. Existuje taktiež mnoho „satirických“ mediálnych organizácií, ktorých existencia je založená na vytváraní týchto nepravdivých článkov (napr. BabylonBee, Onion). Trol potom tieto články zdieľa, prezentuje ich ako pravdivé a keď sa do diskusie zapoja iní užívatelia, trol sa

s týmito používateľmi háda a vytvára spory a medzi ostatnými používateľmi. Trol ale vie, že tieto zdieľané informácie nie sú pravdivé.

Najneprijemnejšie je keď títo trolovia aktívne a opakovane posielajú správy používateľom. Napríklad skupina 10 trolov posielala užívateľovi mnoho súkromných správ ktoré majú väčšinou rovnaký obsah. Sústredia sa na nejakú správu alebo „meme“, ktoré posielajú dookola. Toto sa nazýva „troll swarming“. Trolov vieme rozpoznať väčšinou podľa nasledovných charakteristických znakov [2]:

- Poznámky mimo tému
- Povýšenecký tón a odmietnutie dôkazov
- Používanie nadávok a kontroverzné názory
- Opakujúce sa správy a spam

Existuje množstvo systémov ako rozdeľovať trolov do skupín. Jedno výstižné delenie je znázornené v Obr.1 [3].

Typ skupiny	Typ charakteru trola (orientačne)
Hater (Radi vyvostrujú konverzácie bez žiadneho reálneho benefitu pre ostatných.)	E-Venger (E-Pomstiteľ)
	Iconoclast (Obrazoborec)
	Snert (egoistický neslušný tínedžer)
Lolcow (Radi provokujú ostatných a tým na seba upozorňujú)	Big Man (Veľký muž)
	Ripper (Trhač)
	Chatroom Bob (Četroomový Bob)
Bzzzter (Radi čítajú bez ohľadu na presnosť a užitočnosť svojich vyjadrení)	MHBFY Jenny (Moje srdce pre teba praská Jenny)
	Wizard (Čarodej)
	Flirt (Flirtér)
Oči (Sledujú ako sa vyjadrujú ostatní a čakajú na správnu príležitosť kedy môžu zasiahnuť provokatívnym komentárom)	Lurker (Sliedič)
	Troll (Trol)
	Elder (Staršina)

Obr. 1 Typy a podtypy online trolov.

Existuje viacero osvedčených spôsobov ako trola zastaviť. Najjednoduchšie a najefektívnejšie sú: ignorovať, zapnúť funkciu „mute“, ktorá zastaví notifikácie k jeho príspevkom, zablokovať alebo nahlásiť.

III. NÁVRH PRÍSTUPU K DETEKCIÍ TROLOV

A. Použité regresné metódy

Regresia je štatistický nástroj na odhad vzťahu alebo vzťahov medzi závislou premennou (y alebo výslednou premennou) a jednou alebo viacerými nezávislými premennými (x alebo predikčnou premennou). Uvažovali sme lineárnu, polynomiálnu, logistickú regresiu a Bayesovskú lineárnu regresiu.

Pri lineárnej regresii sa vzťahy modelujú pomocou lineárnych predikčných funkcií, ktorých neznáme parametre modelu sa odhadujú z dát. Takýto model sa nazýva lineárny model [4].

Polynomiálna regresia popisuje polynomiálny vzťah medzi závislou premennou Y (snažíme sa predikovať) a nezávislou premennou X (dostaneme na vstupe). Tento polynóm môže byť n-tého stupňa [5].

Logistickú regresiu je možné použiť na riešenie klasifikačného problému, keďže vďaka použitej sigmoidálnej funkcii transformuje výstupné hodnoty na pravdepodobnosť príslušnosti k triede [6].

Bayesovská lineárna regresia nám umožňuje prístup k užitočnému mechanizmu na vyhodnocovanie nedostatočného počtu údajov alebo zle distribuovaných údajov. Táto metóda nám umožňuje dať prioritu na koeficienty a na šum, takže pri nedostatku údajov koeficienty a šum prednostne prevezmú kontrolu nad fungovaním regresie. Ako sa zvyšuje počet dátových bodov, aposteriórna pravdepodobnosť prepíše apriórnu pravdepodobnosť, čím sa postupne zvyšuje presnosť modelu [7]. Táto regresná metóda je teda veľmi vhodná na odhad hodnoty trolizmu u používateľa, ktorý nám zatiaľ neposkytuje veľa dát. Napríklad, nový používateľ, ktorý ešte toho veľa napísať a zdieľať nestihol.

B. Popis dát

Dáta boli získavané zo sociálnej siete Twitter. Používali sme knižnicu Pandas a Numpy na prácu s tabuľkami a dátami. Knižnicu Matplotlib a Seaborn na vykreslenie grafov pre ľahšiu orientáciu v dátach. Tie obsahovali informácie o 115 používateľoch z 335000 komentárov a príspevkov. Boli stiahnuté dva listy.

Prvý list s názvom Tweet obsahoval: Tweet_ID, Text, Name, Screen_Name, UTC (časová stopa), Created_At, Favorites, Retweets, Language, Client, Tweet_Type, URLs, Hashtag, Mentions, Media_Type a Media, URL.

Druhý list s názvom Users obsahoval: User_ID, Name, Screen_Name, Created_At, Followers, Following, Favorites, Tweets, Lists, Bio, Location, URL, Verified a Default_Profile.

Dátová množina bola labelovaná. Do úvahy sa okrem iného brali vysoká aktivita používateľa, ako je to znázornené na Obr.2 (U bežného používateľa počet príspevkov nepresiahne hodnotu 12) ako aj slovník používateľa. Príklad slovníka trola je ilustrovaný Obr.3.



Obr. 2 Identifikácia trola v dátach. U bežného používateľa počet príspevkov nepresiahne hodnotu 12.

Dáta boli predspracované klasickým spôsobom a výsledná dátová množina obsahovala následné atribúty vrátane labellingu:

- **User** - Meno sledovaného používateľa
- **Avg Post** - Priemerný počet zverejnených tweetov/komentárov za deň
- **Reply_%** - Počet zverejnených komentárov v percentách
- **N** - Počet označení najoznačovanejšieho užívateľa sledovaným užívateľom.
- **1-2** - 1. najoznačovanejší mínus 2. najoznačovanejší užívateľ.
- **Same** - Počet komentárov s rovnakým textom
- **Inapp_%** - Percentuálny podiel správ s nevhodnými slovami.
- **Verified** - Je používateľ Overený? (TRUE/FALSE)
- **Troll** - Je používateľ reálne troll? (TRUE/FALSE)
- **Label** - Aká je hodnota trolizmu tohto užívateľa? (Zistené funkciou na anotovanie užívateľov).

C. Výsledky testov a ich analýza

Návrh prístupu bol implementovaný v Spyder (Anaconda) a v programovacom jazyku Python. Používali sme knižnicu Sklearn na prácu s modelmi regresných metód a program Spyder.

Zo všetkých uvažovaných regresných metód sa najlepšie na osvedčili lineárna regresia a logistická regresia. Výsledky testov modelu lineárnej regresie sú uvedené v Obr. 3. a výsledky logistickej regresie sú v Obr. 4.

Na výpočet výsledkov ukazovateľov miery efektívnosti modelov bol použitý nástroj, ktorý pracuje s hodnotami správne pozitívny (True Positive TP), správne negatívny (True Negative TN), chybné pozitívny (False Positive FP) a chybné negatívny (False Negative FN). Uvedené hodnoty boli vnesené do kontingenčnej tabuľky a z nej boli odvodené hodnoty nasledovných mier efektívnosti Accuracy (Správnosť), Precision (Presnosť), Recall, sensitivity (Návratnosť, Citlivosť), Specificity (Špecifickosť) a F1 Score (Miera F1).

Metrika	Hodnota
Accuracy (Správnosť)	0.8587
Precision (Presnosť)	0.7917
Recall,Sensitivity (Návratnosť, Citlivosť)	0.7037
Specificity (Špecifickosť)	0.9231
F1 Score (Miera F1)	0.7451

Obr. 3 Výsledky testov modelu lineárnej regresie.

Metrika	Hodnota
Accuracy (Správnosť)	0.7283
Precision (Presnosť)	0.3947
Recall,Sensitivity (Návratnosť, Citlivosť)	0.8824
Specificity (Špecifickosť)	0.6933
F1 Score (Miera F1)	0.5455

Obr. 4 Výsledky testov modelu logistickej regresie.

IV. ZÁVER

Článok sa zameriaval na odhaľovanie toxických používateľov typu troll pomocou modelov lineárnej a logistickej regresie. Dôvodom prečo bola lineárna regresia presnejšia je zrejme to, že lineárna regresia pracovala s dátovou množinou so 6 atribútmi a logistická regresia pracovala s dátovou množinou, ktorá bola rozšírená o jeden atribút. Tento atribút sa týkal toho, či je používateľ overený alebo nie. V drvivkej väčšine v množine overených používateľov sa trolovia nenachádzajú, ale naše dáta obsahovali zopár overených užívateľov, ktorí boli trolovia. To mohlo spôsobiť nižšiu presnosť výsledkov logistickej regresie. Z výsledkov testov modelov vyplýva, že zvolenie správnych atribútov a správna úprava dát je jedna veľmi dôležitou súčasťou modelovania, a môže priamo ovplyvniť úspešnosť, efektívnosť a presnosť výsledkov.

POĎAKOVANIE

Práca prezentovaná v tomto článku bola podporená Slovenskou agentúrou na podporu vedy a výskumu v rámci projektu VEGA 1/0685/21.

LITERATÚRA

- [1] DONATH, J. S. (4.8.1996). Identity and deception in the virtual community. In M. A. Smith and P. Kollock (Eds.), *Communities in cyberspace*. London: Routledge. [online][cit 28.4.2022] Dostupné online: https://www.researchgate.net/publication/2512169_Identity_and_Deception_in_the_Virtual_Community.
- [2] What is an Internet Troll? (How to Handle Trolls) <https://www.howtogeek.com/465416/what-is-an-internet-troll-and-how-to-handle-trolls/> [online][cit 10.2.2021].
- [3] BISHOP, Jonathan - Dealing with Internet Trolling in Political Online Communities: Towards the This Is Why We Can't Have Nice Things Scale [online] 2012 [cit 28.4.2022] Dostupné na internete : https://www.researchgate.net/publication/269692808_Dealing_with_Internet_Trolling_in_Political_Online_Communities_Towards_the_This_Is_Why_We_Can't_Have_Nice_Things_Scale.
- [4] HILARY, L. SEAL, X. The historical development of the Gauss linear model. *Biometrika* Vol. 54, Issue 1-2, 1967, pp. 1–24, doi:10.1093/biomet/54.1-2.1. JSTOR 2333849.
- [5] MACHOVA, K., ŠTEFANIK, J. Regression Methods in the Authority Identification within Web Discussions. In: *Computational Collective Intelligence, Lecture Notes in Artificial Intelligence*, Vol. LNAI 9329, no. 1 (2015), © Springer-Verlag Berlin Heidelberg, 2015, 203-212, ISSN 0302-9743.
- [6] MACHOVÁ K. Nové trendy v strojovom učení – Štatistický prístup, Elfa s.r.o., 2016, ISBN: 978-80-553-2602-3.
- [7] WUNDERVALD B. Bayesian Linear Regression [online] 2019 [cit. 28.4.2022] Dostupné online: https://www.researchgate.net/publication/333917874_Bayesian_Linear_Regression>-.

Kryptografický balík Xoodyak a jeho použitie v jednoduchej VPN sieti

¹Marek ROHAČ, ²Miloš DRUTAROVSKÝ

^{1,2} Katedra elektroniky a multimediálnych telekomunikácií, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

¹marek.rohac@student.tuke.sk, ²milos.drutarovsky@tuke.sk

Abstrakt – Článok poukazuje na modernú schému v oblasti tzv. ľahkej (z ang. *lightweight*) kryptografie. Za uvedeným účelom je použitý jeden z finalistov štandardizačného procesu Národného inštitútu pre štandardy a technológie – kryptografický balík Xoodyak. Uvedená kryptografická schéma je opísaná a následne je demonštrovaná v jednoduchej virtuálnej privátnej sieti medzi VPN serverom a klientom. Testovanie bolo realizované vo virtuálnom prostredí operačného systému Linux a voľne dostupného programu DSVPN.

Kľúčové slová – DSVPN, Ľahká kryptografia, Linux, VPN, Xoodyak, XOODOO

I. ÚVOD

Virtuálna privátna sieť (z ang. *Virtual Private Network*, ďalej VPN) [1] je jeden zo spôsobov prepojenia zariadení, tak, že internetová komunikácia medzi nimi je privátna, resp. zabezpečená. To znamená, že komunikácia medzi zariadeniami je šifrovaná a prístup k nej je overený pomocou procesu autentizácie. Vďaka tejto vlastnosti je veľmi výhodné používať danú sieť napríklad keď sa pripájame k internetu pomocou verejnej, nezabezpečenej alebo neznámej siete. Vďaka využitiu tejto technológie je možné ochrániť naše dáta pred zneužitím. V súčasnosti dosahujú VPN siete veľkú popularitu. Verejnosti sú známe napríklad vďaka anonymizácií aktuálnej polohy používateľa.

Podstatou VPN siete je jej privátnosť, resp. zabezpečenie súkromie. Dosiachnutie tejto vlastnosti nám sprostredkuje kryptografia. Konkrétne aplikovaním kryptografických algoritmov a protokolov. Ich cieľom je utajiť dáta pri ich prenose z bodu A do B. Teda od odosielateľa (tvorca) dát, až k ich prijímateľovi. Dôsledkom toho dochádza k zabezpečeniu 3 hlavných úloh:

- **ochrana osobných údajov** (dôvernosť) – z ang. *Data Privacy*,
- **autenticita údajov** (prišli od overeného zdroja) – z ang. *Data Authenticity*,
- **integrita údajov** (neboli upravené počas prenosu) – z ang. *Data Integrity*.

Dôvernosť je najčastejšie žiadaným a známym cieľom. Odosielateľ dát zašifruje ich obsah pomocou použitia niektorého zo šifrovacích algoritmov, kryptografického kľúča a následne dáta odošle. Na druhej strane, prijímateľ, musí použiť komplementárny dešifrovací algoritmus s prislúchajúcim kľúčom. Ktokoľvek, kto sa dostane medzi týchto komunikantov, k takto preposlaným dátam, z nich nedokáže obsahovo nič zistiť. Je potrebné poznamenať, že dáta sú chránené len počas istého časového obdobia. Platí to len do chvíle kým sa útočníkovi nepodarí dešifrovať zašifrované dáta aj bez znalosti tajného kľúča. Tento fakt súvisí s kryptografickou bezpečnosťou použitého šifrovacieho algoritmu [2].

Na základe týchto faktov je jasné, že komunikanti musia mať jasne definovaný použitý kryptografický algoritmus. Zároveň je nutné aby došlo k bezpečnej výmene kryptografických kľúčov, ktoré sú pri šifrovaní a dešifrovaní aplikované. Tým sa zabezpečí aj druhá úloha – autenticita údajov, pretože potrebné informácie budú mať len komunikanti.

V kryptografických algoritmoch sú taktiež aplikované postupy na overenie integrity údajov. Ich úlohou je potvrdiť, že počas transportu nedošlo k žiadnej úprave údajov. Napríklad počas aktívnych útokov na komunikáciu.

Článok sa zameriava na opis kryptografického balíka Xoodyak, ktorý je modernou kryptografickou schémou z oblasti ľahkej kryptografie (z ang. *The lightweight cryptography*). Xoodyak je finalistom štandardizačného procesu [3] Národného inštitútu pre štandardy a technológie (ďalej NIST). V súčasnosti prebieha záverečná fáza štandardizačného procesu NIST [3]. Obsahom článku je aj testovanie použitia uvedenej schémy v jednoduchej VPN sieti medzi serverom a klientom. Kryptografické zabezpečenie je realizované prostredníctvom autentizovaného šifrovania komunikácie medzi komunikantmi. Experimentálne testovanie je realizované vo voľne dostupnom programe DSVPN (z ang. *Dead Simple Virtual Private Network*) vo virtualizovanom operačnom systéme Linux.

II. KRYPTOGRAFICKÝ BALÍK XOODYAK

Tvorcami schémy Xoodyak je tím Keccak, ktorý je známy aj vďaka tvorbe hashovacieho štandardu z rodiny SHA verzie 3 [4]. Xoodyak možno považovať za malý a všestranný kryptografický nástroj, resp. balík. Je vhodný pre väčšinu operácií využívajúcich symetrický kľúč. Napríklad generovanie pseudonáhodných bitov, autentizáciu, šifrovanie, a iné. Zároveň je možné ho použiť aj na hashovania dát. V návrhu Xoodyak bola použitá duplexná konštrukcia s plným stavom a využitím kľúča. Návrh označujeme v angličtine ako *Full-State Keyed Duplex*, (FSKD) [5]. Uvedenú konštrukciu skombinovali s permutáciou XOODOO. Jej opis je obsahom podkapitoly II-A. Pri použití balíka Xoodyak s prednastavenými hodnotami je možné získať 128-bitovú ekvivalentnú kryptografickú bezpečnosť [2].

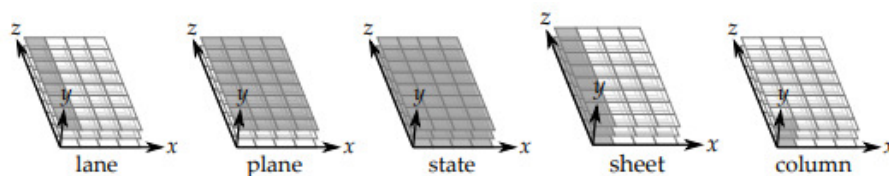
Operačný režim, v ktorom Xoodyak pracuje, sa nazýva Cyklista (z ang. *Cyclist mode*). Tento názov získal ako opak k pomenovaniu režimu Motorista (z ang. *Motorist mode*), ktorý je možné nájsť v Keyak balíku [6], ktorý je predchodcom balíka Xoodyak. Narozdiel od uvedeného balíka Keyak, nie je Xoodyak limitovaný len na autentizované šifrovanie. Viac o režime Cyklista je obsahom podkapitoly II-B.

A. Kryptografická permutácia XOODOO

XOODOO je 384-bitová kryptografická permutácia s parametrizovateľným počtom kôl. Funkcia kola/rundy (z ang. *round*) funguje na 12 slovách (z ang. *words*) po 32 bitoch. Vďaka tomu je efektívna aj na menej výkonných procesoroch. Vytvoril ju obdobne tím Keccak. XOODOO tvorí jadro niektorých kryptografických schém ako napríklad Xoodyak a Xoofff [7].

XOODOO je permutácia, definovaná počtom rúnd. Má klasickú iteračnú štruktúru, ktorá opakovane volá rundovú funkciu s aktuálnym stavom. Operácie v schéme používajú špecifické označenie.

Stav – **state**, pozostáva z 3 rovnako veľkých horizontálnych rovín – **planes**. Každá z týchto rovín obsahuje štyri paralelné 32-bitové pruhy – **lanes**. Okrem tejto charakteristiky je možné opísať stav ako množinu 128 stĺpcov – **columns**, pričom jeden stĺpec obsahuje 3 bity v každej rovine. Stav je teda tvorený zo stĺpcov usporiadaných v poli o rozmere 4×32 . Posledná položka na opis stavu sú tzv. listy – **sheets**. List sa skladá z 3 na sebe uložených pruhov. Uvedené pojmy sú znázornené pomocou obrázku 1, ktorý bol prebraný z [8].



Obr. 1 Grafické znázornenie terminológie využitej v kryptografickej permutácii XOODOO [8]

Roviny majú index y . Index $y = 0$ zodpovedá spodnej rovine a vrchná rovina ma index $y = 2$. Bit je označený s indexom z vrámci množiny pruhov. List označujeme pomocou indexu x . Takže pozícia pruhu v stave je definovaná pomocou dvoch súradníc (x, y) . Konkrétny bit je možné v stave následne reprezentovať pomocou trojice súradníc (x, y, z) . Pri učení stĺpca sú potrebné 2 súradnice (x, z) . Pred spustením samotného permutačného algoritmu musí používateľ vykonať mapovanie 384-bitovej správy voči horizontálnym rovinám. Tento úkon sa realizuje pomocou vzťahu

$$i = z + 32(x + 4y). \quad (1)$$

Rundová funkcia pozostáva z 5 krokov:

- 1) miešanie vrstvy (z ang. *a mixing layer*) θ ,
- 2) posun rovín (z ang. *a plane shifting*) ρ_{west} ,
- 3) prídanie rundových konštánt (z ang. *the addition of round constants*) ι ,
- 4) nelineárna vrstva (z ang. *a non-linear layer*) χ ,
- 5) posun rovín (z ang. *an another plane shifting*) ρ_{east} .

Opis jednotlivých krokov je znázornený pomocou obrázku 4, ktorý je prevzatý z [8]. Obrázok opisuje jednotlivé kroky algoritmu, na ktoré nadväzujú operácie a prvky definované na obrázku 2 a 3. Na obrázku 3 je možné vidieť tabuľku s rundovými konštántami, ktoré sa používajú v algoritme XOODOO permutácie.

A_y	Plane y of state A
$A_y \lll (t, v)$	Cyclic shift of A_y moving bit in (x, z) to position $(x + t, z + v)$
$\overline{A_y}$	Bitwise complement of plane A_y
$A_y + A_{y'}$	Bitwise sum (XOR) of planes A_y and $A_{y'}$
$A_y \cdot A_{y'}$	Bitwise product (AND) of planes A_y and $A_{y'}$

Obr. 2 Definícia prvkov a operácií použitých v algoritme kryptografickej permutácie XOODOO

i	c_i	i	c_i	i	c_i	i	c_i
-11	0x00000058	-8	0x000000D0	-5	0x00000060	-2	0x000000F0
-10	0x00000038	-7	0x00000120	-4	0x0000002C	-1	0x000001A0
-9	0x0000003C0	-6	0x00000014	-3	0x000000380	0	0x00000012

Obr. 3 Rundové konštanty c_i použité v algoritme permutácie XOODOO v hexadecimálnom tvare**Algorithm 1** Definition of XOODOO $[n_r]$ with n_r the number of rounds

Parameters: Number of rounds n_r
for Round index i from $1 - n_r$ to 0 **do**
 $A = R_i(A)$

Here R_i is specified by the following sequence of steps:

θ :

$$P \leftarrow A_0 + A_1 + A_2$$

$$E \leftarrow P \lll (1, 5) + P \lll (1, 14)$$

$$A_y \leftarrow A_y + E \text{ for } y \in \{0, 1, 2\}$$

ρ_{west} :

$$A_1 \leftarrow A_1 \lll (1, 0)$$

$$A_2 \leftarrow A_2 \lll (0, 11)$$

ι :

$$A_0 \leftarrow A_0 + C_i$$

χ :

$$B_0 \leftarrow \overline{A_1} \cdot A_2$$

$$B_1 \leftarrow \overline{A_2} \cdot A_0$$

$$B_2 \leftarrow \overline{A_0} \cdot A_1$$

$$A_y \leftarrow A_y + B_y \text{ for } y \in \{0, 1, 2\}$$

ρ_{east} :

$$A_1 \leftarrow A_1 \lll (0, 1)$$

$$A_2 \leftarrow A_2 \lll (2, 8)$$

Obr. 4 Detailná špecifikácia algoritmu kryptografickej permutácie XOODOO [8]

B. Operačný režim cyklista

Režim cyklista operuje v dvoch režimoch – **hašovací režim a režim kľúča** (z ang. *hash and keyed mode*). Inicializácia prebieha pomocou príkazu `CYCLIST(K, id, counter)`. Ak sa parameter K rovná prázdnomu slovu ϵ , tak potom nastane spustenie v hashovacom režime. Aktuálne nie je do implementácie zakomponovaná možnosť zmeny režimu po inicializácii. Vývojári však túto vlastnosť nevyhlúčili pre prípadné aktualizácie režimu.

Dostupné funkcie závisia od režimu, v ktorom sa režim Cyklista spúšťa. Medzi ne patria `ABSORB()` a `SQUEEZE()`. Možno ich volať v oboch režimoch, zatiaľ čo funkcie `ENCRYPT()`, `DECRYPT()`, `SQUEEZEKEY()` a `RATCHET()` sú dostupné len v režime kľúča. Účel každej funkcie je nasledujúci:

- `ABSORB(X)` absorbuje vstupný reťazec X ,
- $C \leftarrow \text{ENCRYPT}(P)$ zašifruje P do C a absorbuje P ,
- $P \leftarrow \text{DECRYPT}(C)$ dešifruje C do P a absorbuje P ,
- $Y \leftarrow \text{SQUEEZE}(L)$ vytvára L -bajtový výstup, ktorý závisí od doteraz absorbovaných dát,

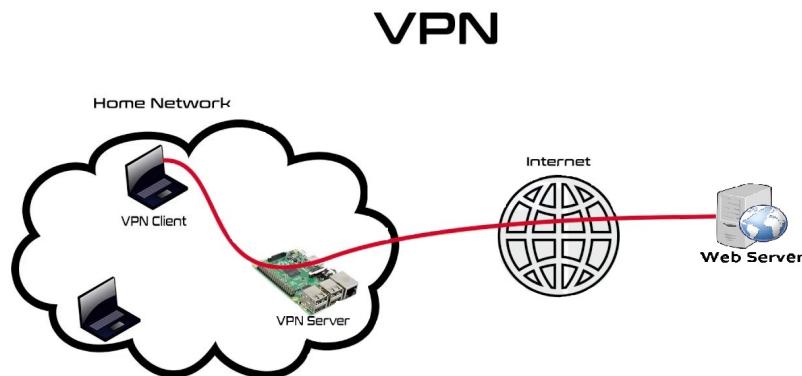
- $Y \leftarrow \text{SQUEEZEKEY}(L)$ funguje ako $\text{SQUEEZE}(L)$, ale používa sa za účelom generovania odvodeného kľúča,
- $\text{RATCHET}()$ transformuje stav na nevratný tak, aby sa zabezpečila dopredná bezpečnosť (z ang. *Forward secrecy*) [9].

Stav bude závisieť od postupnosti volaných funkcií a od jeho vstupných reťazcov. Presnejšie povedané, zámerom je, že akýkoľvek výstup závisí od postupnosti všetkých vstupných reťazcov a volaní, tak, že akékoľvek dva nasledujúce výstupné reťazce budú rôznymi výstupmi. Napríklad volanie $\text{ABSORB}(X)$ znamená, že výstup bude závisieť od reťazca X . Na druhej strane $\text{ABSORB}()$ vo funkcii $\text{ENCRYPT}(P)$ vytvorí výstup závislý aj od P z funkcie šifrovania. Okrem uvedených závislostí ovplyvňujú výstup aj iné dizajnové riešenia. Príkladom je minimalizácia pamäťovej stopy. Vo výsledku teda výstup závisí od počtu predchádzajúcich volaní funkcie $\text{SQUEEZE}()$ a predtým spracovaných textov pomocou funkcií $\text{ENCRYPT}()$ a $\text{DECRYPT}()$ [8].

III. EXPERIMENTÁLNE TESTOVANIE BALÍKA XOODYAK VO VPN SIETI

A. Špecifikácia programu DSVPN

Dead Simple VPN (DSVPN) [10] je voľne dostupný program, napísaný v programovacom jazyku C. Je určený pre operačný systém Linux. Autorom je Frank Denis. DSVPN rieši najbežnejší prípad použitia VPN, teda pripojenie klienta k VPN serveru cez nezabezpečenú sieť. Následne sa klient dostane na internet prostredníctvom servera. Uvedenú skutočnosť je možné vidieť na obrázku 5.



Obr. 5 Prístup k internetu prostredníctvom VPN servera v jednoduchéj VPN sieti

DSVPN používa protokol TCP (z ang. *Transmission Control Protocol*) [11]. Medzi ďalšie výhodné vlastnosti programu patrí:

- Používa iba modernú kryptografiu s formálne overenými implementáciami.
- Malé a konštantné využitie pamäte. Nevykonáva žiadne dynamické alokovanie pamäte (z ang. *heap memory*).
- Malý (približne 25 KB) a čitateľný kód. Žiadne vonkajšie závislosti (z ang. *Dependencies*).
- Funguje po preklade GCC prekladačom. Krátka a prehľadná dokumentácia, žiadny konfiguračný súbor, dodatočná konfigurácia. DSVPN je spustiteľná jednoriadkovým príkazom na serveri, obdobne na klientovi. Bez potreby konfigurácie brány firewall a pravidiel smerovania.
- Funguje na Linuxe (kernel ≥ 3.17), macOS a OpenBSD, DragonFly BSD, FreeBSD a NetBSD v klientskych a point-to-point režimoch. Pridanie podpory pre iné operačné systémy je triviálne.
- Nedochádza k úniku IP adries medzi pripojeniami, ak sa sieť nezmení. Blokuje IPv6 na klientovi, aby sa zabránilo úniku IPv6 adries.

B. Postup pri realizácii experimentu

Xoodyak a DSVPN sme prakticky overili pomocou dvojice virtuálnych strojov (ďalej VM). Pomenovali sme ich Operačný Systém Servera (OSS) a Operačný Systém Klienta (OSC). Na oboch VM sme nainštalovali OS Linux s distribúciou Ubuntu vo verzii 21.03 LTS (z ang. *Long-Term Support*) Na zariadení OSS sme pomocou make [12] a GCC prekladača [13] vykonali inštaláciu DSVPN. Obdobný postup je aplikovaný aj vo VM OSC. Na OSS sme spustili VPN Server,

ktorý nám poskytne IP adresu, prostredníctvom ktorej budeme komunikovať s vonkajším svetom. Autor DSVPN postupoval pri aplikácii schémy Xoodyak presne podľa pokynov vo formálnej dokumentácii [8].

Na spustenie a vytvorenie spojenia sme vykonali nasledujúce úkony:

- 1) Vygenerovanie zdieľaného kľúča:

```
dd if=/dev/urandom of=vpn.key count=1 bs=32
```

– zdieľaný kľúč, ktorý sme vygenerovali, sa nám uložil do súboru *vpn.key*. Ten je potrebné vložiť do priečinka s programom *dsvpn* v oboch zariadeniach – OSS aj OSC.

- 2) OSS zariadenie:

```
sudo ./dsvpn server vpn.key auto
2340 auto 10.8.0.254 10.8.0.2
```

– tento príkaz zabezpečí spustenie VPN servera v prostredí OSS s IP adresou, priradenou k vytvorenému tunelovaciemu rozhraniu s menom *tun0*. Vytvorí sa pri spustení servera¹. Príkazom ďalej definujeme číslo portu 2340, ktoré sa použije pri nadviazaní TCP spojenia medzi klientom a serverom. Poslednou konfiguráciou je priradenie IP adresy tunelov, ktoré bude využívať náš klient – 10.8.0.254 a server – 10.8.0.2. Používateľ má ešte možnosť nastaviť tzv. External IP. Tú by sme využili ak by sme spúšťali DSVPN na smerovači poskytovateľa internetu.

- 3) OSC zariadenie:

```
sudo ./dsvpn client vpn.key 192.168.88.62
2340 auto 10.8.0.2 10.8.0.254
```

– uvedený príkaz zabezpečí, že sa pripojíme na VPN Server, ktorý má IP adresu 192.168.88.62 s portom 2340. Následne sa nadviaže TCP spojenie. Poradie adries tunelov je opačné ako v prípade servera.

- 4) V prípade úspešnej konektivity sme pre okolitý svet viditeľný pomocou IP adresy, ktorú sme nastavili.

Na overenie správnosti funkcionality nám postačí jednoduchý sieťový príkaz *traceroute*. Napríklad *traceroute google.sk*². Prvá z uvedených adries je práve tá, ktorú dané zariadenie používa.

V našom prípade bolo nutné použiť lokálne adresy vzhľadom na to, že oba VM bežia na jednom hosťovskom počítači. Obidve zariadenia sú tým pádom pripojené k jednému internetovému poskytovateľovi, čo má za následok takmer rovnaké smerovanie k vzdialenej internetovej doméne. Je ešte nutné poznamenať, že obom VM sme zmenili nastavenia sieťových adaptérov na možnosť premostenia (z ang. *bridge*). Vďaka čomu im bolo možné priradiť lokálne IP adresy.

Proces zistenia IP adresy VPN servera, po spustení, a overenie funkčnosti je následne znázornené pomocou obrázkov 6, 7.

Nakoniec, na obrázku 7 môžeme vidieť ako klient pri internetovej komunikácii používa namiesto svojej vlastnej, adresu poskytnutú VPN Serverom na zariadení OSS – žltou zvýraznená IP. Červenou je zaškrnutá farba poskytovateľa internetu.

V obrázku 6 sme žltou farbou znázornili IP adresu, na ktorej je VPN server dostupný. Oranžová farba znázorňuje IP adresy tunelu medzi serverom a klientom v tomto poradí. Obdobný proces je možné sledovať aj v prípade klienta vo VM OSC.

IV. ZÁVER

Článok opisuje jeden z trendov v oblasti ľahkej kryptografie. Opisujeme kryptografickú schému Xoodyak vrátane jeho stavebných blokov a jadra – kryptografickej permutácie XOODOO. Uvedená schéma je použitá za účelom zabezpečenia komunikácie medzi klientom a serverom v jednoduchšej VPN sieti. Experimentálne testovanie kryptografickej schémy Xoodyak, je realizované pomocou voľne dostupného a malého programu DSVPN. DSVPN vytvára jednoduchú VPN sieť na základe modelu klient-server, pričom komunikácia VPN klienta s vonkajším svetom je sprostredkovaná pomocou VPN servera. Testovanie je vykonané vo virtualizovanom prostredí OS Linux. Ďalšia podrobnejšia analýza tejto rozsiahlej problematiky využitia kryptografickej permutácie XOODOO, schémy Xoodyak a ich začlenenie do VPN siete bude predmetom nasledujúceho článku.

POĎAKOVANIE

Táto práca bola podporovaná Agentúrou na podporu výskumu a vývoja na základe Zmluvy č. APVV-18-0373.

¹Pomocou `ip address show tun0` zistíme IPv4 adresu VPN servera.

²vo Windows CMD prostredí: `tracert google.sk`


```

ubuntu21@ubuntu21-VirtualBox: ~/Plocha/dsvpn-master
ubuntu21@ubuntu21-VirtualBox: ~/Plocha/dsvpn-master$ ip address show
: lo: <LOOPBACK,UP,LOWER_UP> mtu 65536 qdisc noqueue state UNKNOWN group default qlen 1000
    link/loopback 00:00:00:00:00:00 brd 00:00:00:00:00:00
    inet 127.0.0.1/8 scope host lo
        valid_lft forever preferred_lft forever
    inet6 ::1/128 scope host
        valid_lft forever preferred_lft forever
: enp0s3: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 1500 qdisc fq_codel state UP group default qlen 1000
    link/ether 08:00:27:42:5a:4e brd ff:ff:ff:ff:ff:ff
    inet 192.168.88.60/24 brd 192.168.88.255 scope global dynamic noprefixroute enp0s3
        valid_lft 362sec preferred_lft 362sec
    inet6 fe80::da75:ade1:9470:f56a/64 scope link noprefixroute
        valid_lft forever preferred_lft forever
: tun0: <POINTOPOINT,MULTICAST,NOARP,UP,LOWER_UP> mtu 9000 qdisc fq_codel state UNKNOWN group default qlen 500
    link/none
    inet 10.8.0.254 peer 10.8.0.2/32 scope global tun0
        valid_lft forever preferred_lft forever
    inet6 64:ff9b::a08:fe peer 64:ff9b::a08:2/96 scope global
        valid_lft forever preferred_lft forever
    inet6 fe80::698a:bcc9:aaad:fb43/64 scope link stable-privacy
        valid_lft forever preferred_lft forever
ubuntu21@ubuntu21-VirtualBox: ~/Plocha/dsvpn-master$

```

Obr. 6 Zistenie IP adresy VPN Servera na VM OSS pomocou príkazu ip address show

```

ubuntu20@ubuntu20-VirtualBox: ~/Desktop/dsvpn-master
ubuntu20@ubuntu20-VirtualBox: ~/Desktop/dsvpn-master$ traceroute google.sk
traceroute to google.sk (142.251.36.131), 30 hops max, 60 byte packets
 1 10.8.0.254 (10.8.0.254) 0.963 ms 1.439 ms 1.589 ms
 2 router.lan (192.168.88.1) 2.045 ms 2.205 ms 2.261 ms
 3 192.168.2.222 (192.168.2.222) 2.262 ms 2.686 ms 2.591 ms
 4 192.168.100.1 (192.168.100.1) 8.927 ms 7.779 ms 9.113 ms
 5 172.22.2.33 (172.22.2.33) 9.782 ms 12.546 ms 12.446 ms
 6 172.22.2.1 (172.22.2.1) 18.061 ms 13.177 ms 11.143 ms
 7 172.22.21.1 (172.22.21.1) 9.578 ms 22.255 ms 22.927 ms
 8 * * * 23.712 ms 24.022 ms 22.927 ms
 9 * * *
10 185.171.141.148 (185.171.141.148) 26.491 ms 29.919 ms 29.955 ms
11 185.171.140.8 (185.171.140.8) 29.954 ms 28.328 ms 27.494 ms
12 185.171.140.6 (185.171.140.6) 22.624 ms 22.746 ms 26.569 ms
13 185.171.140.12 (185.171.140.12) 26.734 ms 26.762 ms 26.715 ms
14 185.171.140.254 (185.171.140.254) 27.144 ms 26.904 ms 27.156 ms
15 87.244.236.49 (87.244.236.49) 26.686 ms 26.159 ms 26.180 ms
16 87.244.238.221 (87.244.238.221) 26.132 ms 36.561 ms 29.559 ms
17 87.244.238.78 (87.244.238.78) 24.069 ms 23.925 ms 24.922 ms
18 prg03s12-in-f3.1e100.net (142.251.36.131) 72.807 ms 72.852 ms 72.798 ms
ubuntu20@ubuntu20-VirtualBox: ~/Desktop/dsvpn-master$

```

Obr. 7 Overenie funkcionality DSVPN pomocou príkazu traceroute na VM OSC

LITERATÚRA

- [1] "Virtual private network," [Citované: 18.5.2022]. [Online]. Available: https://en.wikipedia.org/wiki/Virtual_private_network
- [2] "The bit-security of cryptographic primitives," 2017, [Citované: 18.5.2022]. [Online]. Available: <https://wikcr.com/the-bit-security-of-cryptographic-primitives-2/>
- [3] "Lightweight cryptography – finalists," 2022, [Citované: 18.5.2022]. [Online]. Available: <https://csrc.nist.gov/Projects/lightweight-cryptography/finalists>
- [4] "Sha-3," [Citované: 18.5.2022]. [Online]. Available: <https://en.wikipedia.org/wiki/SHA-3>
- [5] J. Daemen, S. Hoffert, S. Mella, M. Peeters, G. Van Assche, and R. Van Keer, "The sponge and duplex constructions," 2020, [Citované: 6.2.2022]. [Online]. Available: https://keccak.team/sponge_duplex.html
- [6] —, "The keyak authenticated encryption scheme," 2020, [Citované: 6.2.2022]. [Online]. Available: <https://keccak.team/keyak.html>
- [7] "Xooff," 2021, [Citované: 6.2.2022]. [Online]. Available: <https://keccak.team/xooff.html>
- [8] J. Daemen, S. Hoffert, S. Mella, M. Peeters, G. Van Assche, and R. Van Keer, "Xoodoo cookbook," 2020, [Citované: 6.2.2022]. [Online]. Available: <https://eprint.iacr.org/2018/767.pdf>
- [9] "Forward secrecy," [Citované: 8.5.2021]. [Online]. Available: https://en.wikipedia.org/wiki/Forward_secretity
- [10] "Dsvpn," [Citované: 18.5.2022]. [Online]. Available: <https://github.com/jedisct1/dsvpn>
- [11] Wikipedia, "Transmission control protocol," [Citované: 26.1.2022]. [Online]. Available: https://en.wikipedia.org/wiki/Transmission_Control_Protocol
- [12] "Gnu make," [Citované: 18.5.2022]. [Online]. Available: <https://www.gnu.org/software/make/>
- [13] "Gcc, the gnu compiler collection," [Citované: 18.5.2022]. [Online]. Available: <https://gcc.gnu.org/>

Softvérovo-definované MANET siete

¹Tomáš HORČÍK, ¹Ján PAPAJ

¹Katedra elektroniky a multimediálnych telekomunikácií, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská Republika

¹tomas.horcik@student.tuke.sk, ¹jan.papaj@tuke.sk

Abstract – Softvérovo-definované siete a ich benefity predstavujú obrovský krok vpred pre súčasné mobilné siete, ale aj pre siete ďalších generácií. Spojenie softvérovo-definovaného prístupu s mobilnými ad-hoc sieťami predstavuje nový spôsob, ako čo najlepšie využiť dostupné zdroje týchto sietí. Okrem výhod však tento nový spôsob riadenia mobilných ad-hoc sietí prináša aj určité problémy, ktoré je potrebné riešiť. Cieľom tejto práce je poukázať na nové možnosti smerovania v softvérovo-definovaných mobilných ad-hoc sieťach a ich využitie v prostredí simulátora Mininet-WiFi.

Keywords – algoritmus, MANET, SD-MANET, Mininet-WiFi, smerovanie

I. ÚVOD

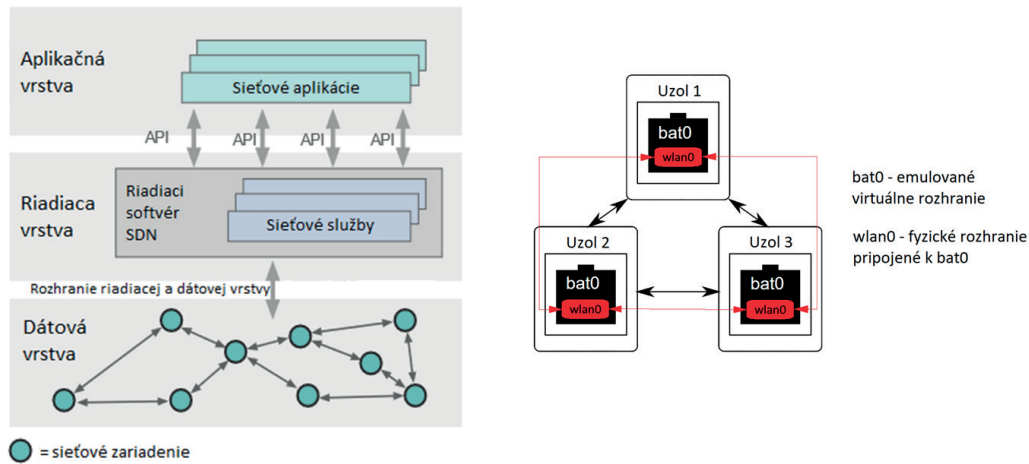
Softvérovo-definované siete (SDN) predstavujú o ich spojenie s mobilnými ad-hoc sieťami (MANET) predstavuje nový spôsob dátovej komunikácie s ohľadom na čo najlepšie využitie dostupných komunikačných zdrojov [1]. Okrem výhod však tento nový spôsob riadenia MANET prináša aj určité problémy, ktoré je potrebné riešiť. Cieľom tejto publikácie je poukázať, akými spôsobmi je možné realizovať smerovanie v softvérovo-definovaných MANET (SD-MANET), ktoré sú schopné umožniť komunikáciu medzi jednotlivými komponentami sieťovej architektúry takejto siete [2]. Ďalej to je poskytnutie praktickej implementácie SD-MANET v simulačnom prostredí, ktoré môže byť nápomocné pri riešení existujúcich problémov.

Hlavnou myšlienkou SDN je oddelenie a logické centralizovanie riadiacej časti v softvérovo-definovanej riadiacej jednotke (ďalej len riadiaca jednotka) (Obr. 1) [3]. Tento prístup je založený na virtuálizácii a umožní tak sieťovým zariadeniam venovať všetok svoj procesný výkon iba na rozposielanie dát, čím sa tento proces výrazne zrýchli. Ďalšou výhodou prítomnosti centralizovanej logiky je manažment siete. Bežná sieť na konfiguráciu počiatočných nastavení alebo zmeny konfigurácie vyžaduje individuálny fyzický prístup k zariadeniam [4]. V SDN sa zmeny v konfigurácii môžu aplikovať na zariadenia z jedného miesta, čo ušetrí sieťovému administrátorovi čas a predíde potenciálnym chybným konfiguráciám. Medzi najdôležitejšie aspekty siete patrí predovšetkým bezpečnosť. Riadiaca jednotka dokáže poskytnúť vysokú úroveň bezpečnosti pre celú sieť aplikovaním správnych bezpečnostných pravidiel. Taktiež, šírenie sieťovej infekcie môže byť zastavené administrátorom z jedného centrálného bodu. Ďalšou veľmi dôležitou výhodou softvérovo-definovaného prístupu je celkové zníženie nákladov v sieti. Konkrétne sa jedná napríklad o automatizovanú administráciu siete, nakoľko bežné úlohy a problémy spojené s ňou môžu byť automatizované, alebo sa môže jednať o efektívne využitie serverových zdrojov. Jednou z najväčších výhod softvérovo-definovaného prístupu je možnosť manipulovať sieťovú trafiku, čo nám umožňuje dosiahnuť požadovanú kvalitu služieb (QoS) v sieti [5].

Tak ako každá technológia, aj SDN so sebou prinášajú svoje nevýhody. Integrovanie softvérovo-definovaného protokolu a riadiacej jednotky vyžaduje zmenu v celej sieťovej infraštruktúre. Kvôli tejto zmene vzniká aj potreba kompletnej zmeny v konfigurácii siete, ktorá tak môže zvýšiť celkové náklady. Zamestnanci musia byť príslušne vyškolení a musia sa takisto naučiť používať nové nástroje na manažment siete. Bezpečnosť v takejto sieti predstavuje ťažkú úlohu, ktorú zvládnu iba experti na danú problematiku, pretože zatiaľ neexistuje žiadny štandardizovaný bezpečnostný protokol, ktorý by sa o to jednoducho postaral [5].

A. MANET a softvérovo-definované siete

Spontánna konfigurácia a heterogénnosť MANET je citlivá na efektívne využívanie komunikačných zdrojov. Aby bola zaručená správna kvalita služieb (QoS), je potrebné zabezpečiť vzájomnú spoluprácu uzlov v sieti [6], [7]. Softvérovo-definovaný prístup toto obmedzenie dokáže vyriešiť vďaka schopnosti udržiavať si globálny nadhľad siete, ktorý výrazne obohatí smerovacie rozhodnutia o informácie potrebné na zvolenie optimálnej cesty. Takýto nadhľad celej topológie



Obr. 1: Základná architektúra SDN a komunikácia medzi rozhraniami

umožňuje riadiacej jednotke zohľadňovať rôzne úrovne priority dátového prenosu a adekvátne tomu prispôbiť výber cesty. Môže nastať situácia, kedy v MANET budú chcieť dva komunikačné uzly využiť rovnakú dátovú cestu, pretože sa bude zdať ako cenovo najlepšia. Ani jeden z daných dátových prúdov nevyužije alternatívnu cestu, pretože bude zdanlivo horšia ako tá hlavná cesta. Prítomnosť riadiacej jednotky túto situáciu dokáže vyriešiť tým, že jednému prúdu prideliť alternatívnu cestu. Ak by nastal prípad, že alternatívna cesta neexistuje, dátovému prúdu s nižšou prioritou budú pridelené redukované zdroje siete po celý čas prenosu dátového prúdu s vyššou prioritou [2], [7]. SD-MANET bez priamo spojenia každého uzla s riadiacou jednotkou pomocou separátneho komunikačného kanálu a lokalizačnej služby na zistenie polohy uzlov potrebuje komunikovať s riadiacou jednotkou metódou viacpreskokových ciest. Riadiaca jednotka musí v takomto prostredí [2]:

- Posielať aktualizácie všetkým uzlom v sieti ohľadom každej zmeny v topológii.
- Prieběžne od uzlov zbierať informácie o stave konektivity a učiť sa sieťovú topológiu.
- Udržovať si cestu ku každému uzlu.

Na splnenie týchto požiadaviek využíva riadiaca jednotka 3 interné moduly [2]:

1) *Manažér konektivity (Connectivity Manager CM)*

- Topológia siete sa často mení a uzly v sieti si potrebujú udržiavať cestu k riadiacej jednotke, aby od nej mohli dostávať riadiace správy.
- CM napomáha uzlom nájsť a udržať si túto cestu k riadiacej jednotke.

2) *Manažér topológie (Topology Manager TM)*

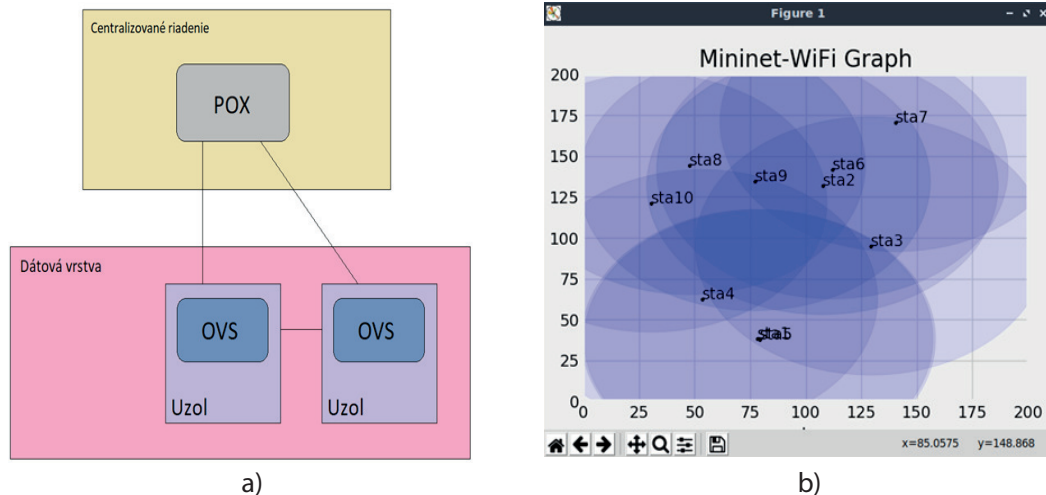
- Riadiaca jednotka potrebuje prieběžne zisťovať, ako vyzerá celá topológia siete, aký je momentálny počet uzlov a v akom stave sú jednotlivé spojenia medzi nimi.
- TM je zodpovedný za zbieranie a udržiavanie týchto informácií.

3) *Manažér preposielania (Forwarding Manager FM)*

- Pomocou tohto modulu riadiaca jednotka posielala sieťové cesty jednotlivým uzlom v sieti.

Aj napriek všetkým existujúcim výzvam v MANET je implementovanie softvérovo-definovanej architektúry obrovským prínosom. Proces zisťovania topológie a výberu smerovacích ciest je presunutý na riadiacu jednotku, čo výrazne zníži spotrebu vstavanej batérie mobilných zariadení [8]. Keďže sa smerovacie informácie nemusia vymieňať priamo medzi susedmi, sú vďaka tomu znížené režijné náklady na komunikáciu. Výkon siete je zvýšený v dôsledku dynamického upravovania smerovacích parametrov v sieti na základe momentálnej sieťovej charakteristiky. Oneskorenie v sieti je redukované vďaka oportunistickému aktualizovaniu smerovacích ciest na základe požiadaviek alebo zmien v topológii. V prípade, že je dostupný pohybový model pre danú sieť je možné dosiahnuť bezproblémovú a nerušnú komunikáciu preventívnym posielaním smerovacích informácií. Uzly, ktoré majú viacero rádiových prijímačov nakonfigurovaných na rôzne frekvencie umožňujú riadiacej jednotke efektívne využívať rádiové spektrum a zvýšiť tak výkon siete [7], [8].

Aplikovanie SD-MANET predstavuje aj určité komplikácie, ktoré je potrebné vyriešiť. Ako už bolo spomenuté, MANET sa vyznačujú vysoko dynamickým pohybom jednotlivých uzlov, čo spôsobuje, že topológia takejto siete nie je organizovaná, resp. je neštruktúrovaná. Jednotlivé uzly nemajú vytvorené oddelené komunikačné kanály s riadiacou jednotkou, a preto je nutné



Obr. 2: SD-MANET: a) Vytvorená architektúra SD-MANET, b) Simulačný scenár v simulátore MiniNET-WiFi

aby si každý uzol v sieti bol schopný k nej nájsť cestu. Uzly v mobilnej ad-hoc sieti majú obmedzený rádiový dosah, čo môže predstavovať problém pri osobitnom naviazaní kontaktu s riadiacou jednotkou. Musia teda komunikovať na viac preskokov, čo môže spôsobovať nespoľahlivé spojenie. Objavujú sa časté kolízie paketov a vysoká interferencia pri súčasnom vysielaní, ktoré sú spôsobené nízkou kapacitou spojení a nesynchronizovaným prenosom [9].

II. IMPLEMENTÁCIA SD-MANET V SIMULÁTORE MININET - WiFi

Navrhnutá schéma SD-MANET je zobrazená na obr. 2a, kde je jasne vidieť rozdelenie siete na dve vrstvy. Na vrchnej riadiacej vrstve sa nachádza riadiaca jednotka *POX*, ktorá je individuálne pripojená na každý uzol v sieti. Vyplýva z toho teda, že táto implementácia využíva centralizované riadenie. Na dolnej dátovej vrstve sa nachádzajú mobilné zariadenia, resp. uzly, na ktorých beží *OVS* softvér. V implementácii je možné použiť ľubovoľnú dostupnú riadiacu jednotku.

Vytvorenie SD-MANET spočíva aj v úprave MANET tak, aby bola schopná spolupracovať s riadiacou jednotkou. Simulačný nástroj *Mininet-WiFi* však neposkytuje možnosť pripojenia riadiacej jednotky s mobilnými zariadeniami, čo predstavuje veľký problém. Preto bol navrhnutý model mobilného zariadenia, ktoré sa zároveň správa aj ako *OpenFlow* prepínač [10]. *Mininet-WiFi* poskytuje rozšírenie zvané *Containernet* [11], ktoré umožňuje mobilné zariadenia vytvárať z *Docker* kontajnerov [12]. Toto vylepšenie tak poskytuje možnosť nainštalovania hocikákeho softvéru na mobilné zariadenia. Na vytvorenie mobilného *OpenFlow* prepínača som použil *Docker* obraz, ktorý obsahuje softvérový prepínač *Open vSwitch* (*OVS*) [13]. Týmto postupom sme získali prototyp úmobilného zariadenia, na ktorom môže bežať inštancia *OpenFlow* prepínača.

Na správne fungovanie v SD-MANET však bolo potrebné vykonať niekoľko úprav pôvodného *Docker* obrazu. Pôvodný *Docker* obraz bol založený na operačnom systéme *Ubuntu* verzie 16.04, ktorý však nefungoval správne, a preto bol aktualizovaný na verziu 18.04. Aby boli zariadenia v sieti schopné medzi sebou komunikovať, bolo potrebné do obrazu nainštalovať vybraný smerovací protokol *Batman-advanced* a nástroj *ping* pre možnosť testovania komunikácie medzi zariadeniami [9]. Okrem spomenutých dodatkov bolo potrebné nainštalovať aj nástroj *batctl*, ktorý slúži na spravovanie a riadenie zariadení využívajúcich smerovací protokol *Batman-advanced*. Keďže tento protokol pracuje na druhej vrstve *ISO/OSI* modelu, zariadenia sú pre protokoly vyšších vrstiev transparentné, a preto bežné diagnostické nástroje nefungujú tak, ako by mali. Po nainštalovaní všetkých potrebných závislostí do pôvodného *Docker* obrazu bolo ďalej potrebné vyriešiť vnútornú architektúru mobilného zariadenia, ktoré nefungovalo správne kvôli neschopnosti smerovacieho protokolu spolupracovať s inštanciou *OpenFlow* prepínača.

Veľkým problémom bolo, že rozhodnutia smerovacieho protokolu nebolo možné ovplyvniť pravidlami v tabuľkách tokov v *OpenFlow* prepínači. Prichádzajúce pakety boli smerované iba na základe smerovacej tabuľky protokolu *Batman-advanced*. Tento problém bolo možné vyriešiť tak, že rozhranie *bat0* (Obr. 3), ktoré je automaticky vytvorené na zariadení protokolom *Batman-advanced*, bolo pridané medzi rozhrania na *OpenFlow* prepínači. Týmto spôsobom sa všetky pakety okrem smerovacej tabuľky porovnávali aj so vstupmi v tabuľke tokov v *OpenFlow* prepínači.

Po dosiahnutí úspešnej spolupráce protokolu *Batman-advanced* a *OpenFlow* prepínača bolo následne možné pripojiť riadiacu jednotku k zariadeniu. Táto implementácia využíva riadiacu jednotku *POX* (Obr. 3, ktorá je založená na takzvaných komponentoch. Tieto komponenty poskytujú


```
root@sta1:~# ifconfig
bat0: flags=4163<UP,BROADCAST,RUNNING,MULTICAST> mtu 1500
    inet 192.168.123.1 netmask 255.255.255.0 broadcast 0.0.0.0
    ether 2a:f0:11:73:b5:86 txqueuelen 1000 (Ethernet)
    RX packets 0 bytes 0 (0.0 B)
    RX errors 0 dropped 0 overruns 0 frame 0
    TX packets 0 bytes 0 (0.0 B)
    TX errors 0 dropped 0 overruns 0 carrier 0 collisions 0

eth0: flags=4163<UP,BROADCAST,RUNNING,MULTICAST> mtu 1500
    inet 172.17.0.2 netmask 255.255.0.0 broadcast 172.17.255.255
    ether 02:42:ac:11:00:02 txqueuelen 0 (Ethernet)
    RX packets 36 bytes 4110 (4.1 KB)
    RX errors 0 dropped 0 overruns 0 frame 0
    TX packets 0 bytes 0 (0.0 B)
    TX errors 0 dropped 0 overruns 0 carrier 0 collisions 0

lo: flags=73<UP,LOOPBACK,RUNNING> mtu 65536
    inet 127.0.0.1 netmask 255.0.0.0
    loop txqueuelen 1000 (Local Loopback)
    RX packets 0 bytes 0 (0.0 B)
    RX errors 0 dropped 0 overruns 0 frame 0
    TX packets 0 bytes 0 (0.0 B)
    TX errors 0 dropped 0 overruns 0 carrier 0 collisions 0

sta1-wlan0: flags=4099<UP,BROADCAST,MULTICAST> mtu 1500
    inet 10.0.0.3 netmask 255.0.0.0 broadcast 10.255.255.255
    ether 00:02:00:00:00:10 txqueuelen 1000 (Ethernet)
    RX packets 32350 bytes 2402860 (2.4 MB)
    RX errors 0 dropped 0 overruns 0 frame 0
    TX packets 6503 bytes 623982 (623.9 KB)
    TX errors 0 dropped 0 overruns 0 carrier 0 collisions 0
```

Obr. 3: Ukážka rozhrania bat0 na mobilnom zariadení v SD-MANET

```
root@sta1:~# ovs-vsctl show
94071ed2-3b93-47fb-8f60-aef692722e09
    Bridge "br0"
        Controller "tcp:192.168.56.107:6633"
        Port "br0"
            Interface "br0"
                type: internal
        Port "bat0"
            Interface "bat0"
    ovs_version: "2.11.2"
```

a)

```
wifi@wifi-virtualbox:~/pox$ ./pox.py forwarding.l2_learning
POX 0.7.0 (gar) / Copyright 2011-2020 James McCauley, et al.
WARNING:version:Support for Python 3 is experimental.
INFO:core:POX 0.7.0 (gar) is up.
```

b)

Obr. 4: Ukážka činnosti SD-MANET: a) Rozhranie bat0 po pridaní na OpenFlow, b) Riadiaca jednotka POX s komponentom forwarding.l2_learning

rôzne funkcionality, ktoré je možné si naprogramovať podľa vlastných potrieb. Komponent využitý pri simulovaní SD-MANET sa nazýva *forwarding.l2_learning* (Obr. 4 b). Tento komponent slúži iba na demonštráciu a dokáže prinútiť OpenFlow prepínač správať sa ako obyčajný prepínač pracujúci na druhej vrstve ISO/OSI modelu.

III. SIMULÁCIA A VÝSLEDKY

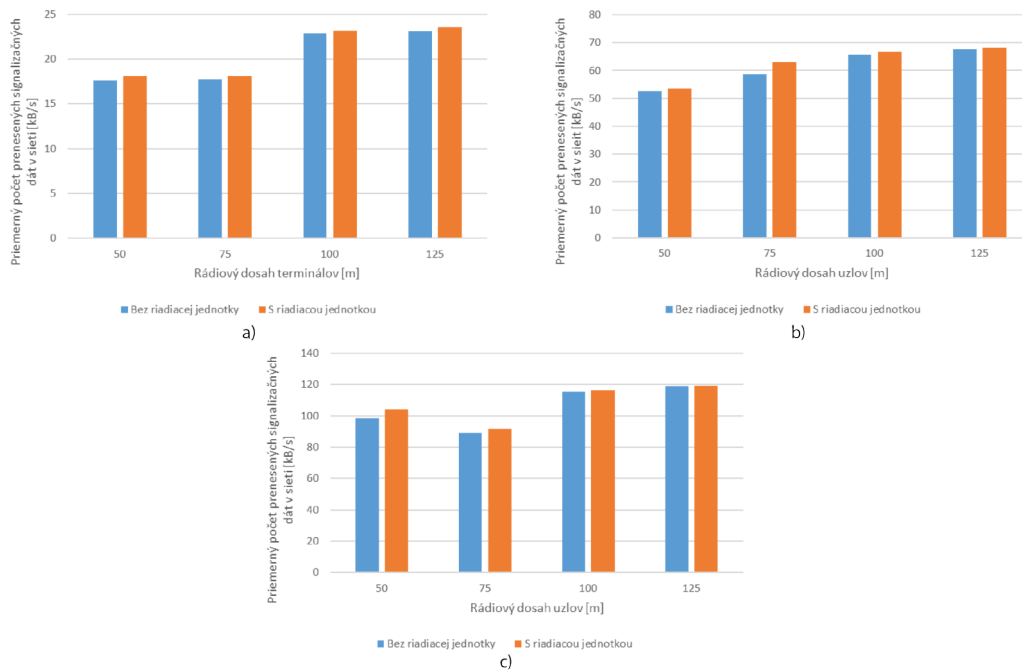
Simulačný experiment pozostával z dvoch scenárov. V prvom scenári sme sa zamerali na analyzovanie činnosti siete z pohľadu počtu signalizačných paketov v sieti a teda boli analyzované *priemerné hodnoty prenesených signalizačných dát*[kB/s] a *percentuálne zastúpenie paketov určených na smerovanie* v SD-MANET. Druhý experiment bol zameraný na analyzovanie činnosti smerovacieho protokolu *Batman-advance* s implementovanou SD technológiou. Siete s počtom mobilných uzlov 10, 20 a 30 sa testovali s pripojenou a bez pripojenej riadiacej jednotky. Každý jeden z týchto scenárov sa ešte k tomu testoval s rôznymi nastaveniami rádiového dosahu mobilných zariadení (50, 75, 100, 125 m). Zber dát v simulovanej sieti sa vykonával po dobu 100 sekúnd s počtom opakovaní 10. Tab. 1 zobrazuje sieťové parametre použité v simulačnom experimente.

A. Experiment 1

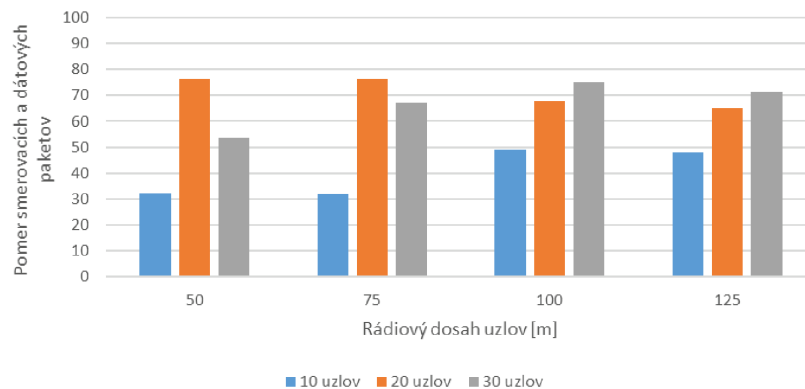
Získavanie dát bolo vykonávané počas činnosti riadiacej jednotky s rôznymi rádiovými dosahmi mobilných uzlov v SD-MANET. Výsledky v tejto kategórii naznačujú, že väčší rádiový dosah uzlov má za následok zvýšenie počtu priemerných prenesených dát [kB/s] v SD-MANET. Jedinou výnimkou je 75 metrový rádiový dosah mobilných uzlov v sieti s počtom mobilných uzlov

Parameter	Hodnota
Počet uzlov	10, 20, 30
Simulačná plocha [m ²]	200 x 200
Čas simulácie [s]	100
Model šírenia	model long-distance
Rádiový dosah terminálov [m]	50, 75, 100, 125
Smerovací protokol	Batman-advance
Model pohybu	Náhodný model pohybu
Počet opakovaní	10

Tabuľka 1: Parametre simulácie v simulátore Mininet-WiFi



Obr. 5: Závislosť počtu prenesených signalizačných dát [kB/s] od rádiového dosahu pre: a) 10 uzlov, b) 20 uzlov, c) 30 uzlov



Obr. 6: Podiel smerovacích paketov protokolu Batman-advanced v SD-MANET [%]

30, ktorý vykazuje náhly pokles v nameraných hodnotách (Obr. 5). Meranie bolo vykonané v prítomnosti riadiacej jednotky s rôznymi hodnotami rádiových dosahov uzlov v sieti. Smerovacie pakety tvoria veľkú časť všetkých paketov používaných v SD-MANET, pretože slúžia na periodické informovanie ostatných uzlov v sieti o existencii pôvodného uzla.

B. Experiment 2

Meranie bolo vykonané v prítomnosti riadiacej jednotky s rôznymi hodnotami rádiových dosahov uzlov v sieti. Smerovacie pakety tvoria veľkú časť všetkých paketov používaných v SD-

MANET, pretože slúžia na periodické informovanie ostatných uzlov v sieti o existencii pôvodného uzla (Obr. 6). Na základe dosiahnutých výsledkov môžeme konštatovať, že samotná činnosť SD-MANET neovplyvní dátovú komunikáciu redundantnými paketmi, ale iba dátami, ktoré sú nevyhnutné k činnosti SD-MANET.

IV. ZÁVER

SD-MANET je nový koncept mobilných viacpreskokových sietí, ktoré budú implementované do prostredia 6G sietí a budú využívané na komunikáciu nielen v krízových situáciách. Softvérovo-definovaný prístup predstavuje významný krok vpred vo svete počítačových sietí, kde tento nový spôsob riadenia siete zaznamenal pozitívne ohlasy. Aplikovať však tento nový model do MANET sietí predstavuje radu problémov a obmedzení, ktoré je potrebné postupne vyriešiť. Z dosiahnutých výsledkov je zrejmé, že pri využívaní SD-MANET sa musíme zamerať aj na samotnú optimalizáciu smerovacích procesov.

POĎAKOVANIE

Táto publikácia vznikla vďaka podpore v rámci Operačného programu Integrovaná infraštruktúra pre projekt: Výskum v sieti SANET a možnosti jej ďalšieho využitia a rozvoja, kód ITMS: NFP313010W988, spolufinancovaný zo zdrojov Európskeho fondu regionálneho rozvoja.

LITERATÚRA

- [1] M. Rahouti, K. Xiong, Y. Xin, S. K. Jagatheesaperumal, M. Ayyash, and M. Shaheed, "Sdn security review: Threat taxonomy, implications, and open challenges," *IEEE Access*, vol. 10, pp. 45 820–45 854, 2022.
- [2] A. Dusia and A. S. Sethi, "Software-defined architecture for infrastructure-less mobile ad hoc networks," in *2021 IFIP/IEEE International Symposium on Integrated Network Management (IM)*, 2021, pp. 742–747.
- [3] D. Kafetzis, S. Vassilaras, G. Vardoulas, and I. Koutsopoulos, "Software-defined networking meets software-defined radio in mobile ad hoc networks: State of the art and future directions," *IEEE Access*, vol. 10, pp. 9989–10014, 2022.
- [4] R. Gokilabharathi and P. Deepalakshmi, "Efficient load balancing to enhance the quality of service (qos) in software defined networking (sdn)," in *2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI)*, 2018, pp. 1046–1051.
- [5] U. Srilakshmi, S. A. Alghamdi, V. A. Vuyyuru, N. Veeraiah, and Y. Alotaibi, "A secure optimization routing algorithm for mobile ad hoc networks," *IEEE Access*, vol. 10, pp. 14 260–14 269, 2022.
- [6] H. Zhang, X. Wang, P. Memarmoshrefi, and D. Hogrefe, "A survey of ant colony optimization based routing protocols for mobile ad hoc networks," *IEEE Access*, vol. 5, pp. 24 139–24 161, 2017.
- [7] C. Giannelli, P. Bellavista, and D. Scotece, "Software defined networking for quality-aware management of multi-hop spontaneous networks," in *2018 International Conference on Computing, Networking and Communications (ICNC)*, 2018, pp. 561–566.
- [8] K. Kiran, N. P. Kaushik, S. Sharath, P. D. Shenoy, K. R. Venugopal, and V. T. Prabhu, "Experimental evaluation of batman and batman-adv routing protocols in a mobile testbed," in *TENCON 2018 - 2018 IEEE Region 10 Conference*, 2018, pp. 1538–1543.
- [9] L. Barolli, M. Ikeda, G. De Marco, A. Durresi, and F. Xhafa, "Performance analysis of olsr and batman protocols considering link quality parameter," in *2009 International Conference on Advanced Information Networking and Applications*, 2009, pp. 307–314.
- [10] S. Li, Y. Shao, S. Ma, N. Xue, S. Li, D. Hu, and Z. Zhu, "Flexible traffic engineering: When openflow meets multi-protocol ip-forwarding," *IEEE Communications Letters*, vol. 18, no. 10, pp. 1699–1702, 2014.
- [11] R. R. Fontes, S. Afzal, S. H. B. Brito, M. A. S. Santos, and C. E. Rothenberg, "Mininet-wifi: Emulating software-defined wireless networks," in *2015 11th International Conference on Network and Service Management (CNSM)*, 2015, pp. 384–389.
- [12] S. Singh and N. Singh, "Containers amp; docker: Emerging roles amp; future of cloud technology," in *2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT)*, 2016, pp. 804–807.
- [13] P. Krongbaramee and Y. Somchit, "Implementation of sdn stateful firewall on data plane using open vswitch," in *2018 15th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, 2018, pp. 1–5.

Detekcia kvality orezu moriacich liniek

¹Dominik TKÁČIK, ²Csaba SZABÓ, ³Róbert VANDLÍK

^{1,2}Katedra počítačov a informatiky, Fakulta elektrotechniky a informatiky, Technická
Univerzita v Košiciach, Slovenská republika

³U.S.Steel Košice a.s., Slovenská republika

¹tkacik38@gmail.com, ²csaba.szabo@tuke.sk, ³RVandlik@sk.uss.com

Abstrakt – Článok je zameraný na vytvorenie riešenia pre automatizovanú evaluáciu kvality orezu oceľového pásu na moriacej linke s využitím počítačového videnia metódou neurónových sietí. Hlavným cieľom výskumu je dosiahnutie ekonomickej úspory pomocou včasného upozornenia operátora výroby na zhoršenú kvalitu orezu oceľového pásu. Riešenie je nasadené v reálnom procese výroby na moriacich linkách v spoločnosti U.S.Steel Košice, kde je operátor výroby informovaný o zhoršenej kvalite orezu prostredníctvom vizuálneho a zvukového alarmu. V samotnom riešení sme aplikovali architektúru neurónovej siete MLP mixer na vyriešenie problému malého počtu dôležitých pixelov pre určenie kvality orezu.

Kľúčové slová – Hlboké učenie, Konvolučné neuronové siete, MLP mixer, Moriaca linka, Optimalizácia výrobného procesu

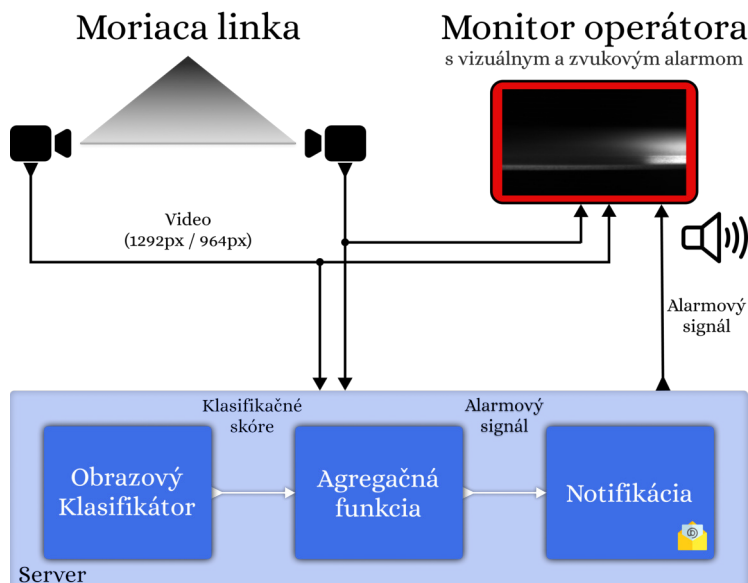
I. ÚVOD

Tento článok je venovaný detekcií zlého orezu na moriacich linkách a automatizácií kontroly orezu. Viedol nás k tomu najmä fakt, že kvôli zlému orezu stráca oceľový zvitok na svojej hodnote a je potrebné ďalšie spracovanie, ktoré navyšuje výrobné náklady [1], [2]. V prípade ak nie je možné oceľový zvitok dodatočne upraviť, zvitok sa stáva nepredajným a je potrebné ho zrecyklovať, čím sa vráti na začiatok výrobného procesu. V súčasnej situácii je kontrola kvality orezu vykonávaná operátorom moriacej linky prostredníctvom snímok orezov. Náplň jeho práce spočíva v obsluhu moriacej linky a kontrola kvality orezu predstavuje len minoritnú časť jeho práce. Problematické je samotné rozpoznanie nekvalitného orezu oceľového pásu, nakoľko rozdiel medzi dobrým a zlým orezom je reprezentovaný len niekoľkými pixelmi na snímke orezu. Z tohto dôvodu je vysoká pravdepodobnosť neskorého odhalenia zlého orezu a teda aj znehodnotenia samotného oceľového zvitku.

V súčasnej dobe neexistuje žiadne technologické riešenie pre automatizáciu kontroly orezu a teda, zlepšenie kvality orezu by bolo možné vykonať len navýšením personálu, ktorý by prioritne kontroloval kvalitu samotného orezu oceľového pásu. Podobný výskum sa nachádza v [3], avšak pre iné technológie. Cieľom nášho výskumu je navrhnutie a implementovanie technologického riešenia, ktoré bude schopné znížiť vplyv ľudského faktora na vyhodnocovanie kvality orezu. Nami navrhované riešenie by malo byť schopné prostredníctvom kamier detekovať a včas informovať operátora o zlej kvalite orezu oceľového pásu. Finálne riešenie bude aplikované na moriacej linke v spoločnosti U.S.Steel Košice a jeho funkčnosť bude overená v reálnych podmienkach. Riešenie musí byť schopné spoľahlivo určiť kvalitu orezu aj v prípade zvlnenia oceľového pásu, rôznych širok a hrúbok oceľového pásu a v neposlednom rade, musí byť aplikovateľné na akúkoľvek moriacu linku. Riešenie bude využívať metódu počítačového videnia [4], [5], [6] a snímky z existujúceho kamerového systému určeného na snímanie orezu pre manuálnu kontrolu jej kvality operátorom.

II. KONCEPTUÁLNY NÁVRH RIEŠENIA

Na obrázku 1 môžeme vidieť konceptuálny návrh riešenia nášho problému. Naše riešenie bude sledovať snímky z kamier a využitím techník počítačového videnia bude klasifikovať kvalitu orezu na jednotlivých snímkach. Taktiež je potrebné vytvoriť agregáčnú funkciu, ktorá na základe sledu predchádzajúcich pravdepodobností klasifikačného modelu v rozmedzí prahových hodnôt bude vyhodnocovať aktiváciu alarmu. V prípade, že sa alarm aktivuje, je potrebné o tejto skutočnosti informovať vedúceho prevádzky a zároveň spracovať alarmový signál na operátorovom monitore. Pri aktivovaní alarmu sa operátorov monitor rozbliká na červeno, pričom blikanie bude sprevádzané zvukovým signálom.



Obr. 1 Konceptuálny návrh riešenia.

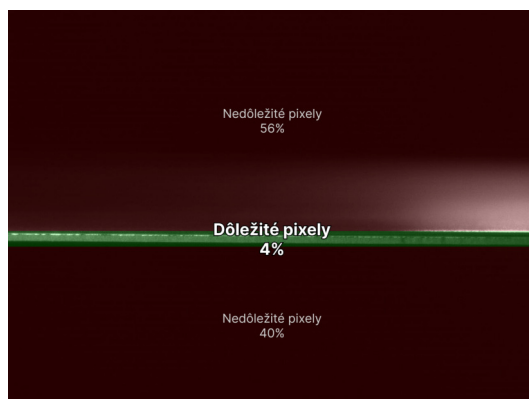
III. OBRAZOVÝ KLASIFIKÁTOR

Pre klasifikáciu snímok orezu oceleovej hrany sme natrénovali model pomocou techník strojového učenia s učiteľom. Tento model na základe snímky na vstupe je schopný klasifikovať kvalitu orezu oceleovej hrany do troch kategórií a to na *dobrý orez*, *zlý orez* a *nedá sa určiť*. Posledná klasifikačná trieda nazvaná *nedá sa určiť*, je podstatná z dôvodu že môže nastať hardvérová porucha, ktorá spôsobí znehodnotenie snímky v podobe rozostrenia alebo nulovej viditeľnosti.

A. Snímka orezu

Naše snímky orezov z kamery sú vo veľkosti 1292 x 962 pixelov. Z celkovej výšky obrázka sa hrana plechu vyskytuje len zhruba v 4% obrázka, čo predstavuje zhruba 30 pixelov vysoký pás. Výška oceleovej hrany na obrázku je závislá od hrúbky a šírky oceleového pásu. Práve šírka oceleového pásu určuje vzdialenosť medzi snímanou hranou pásu a objektívom kamery.

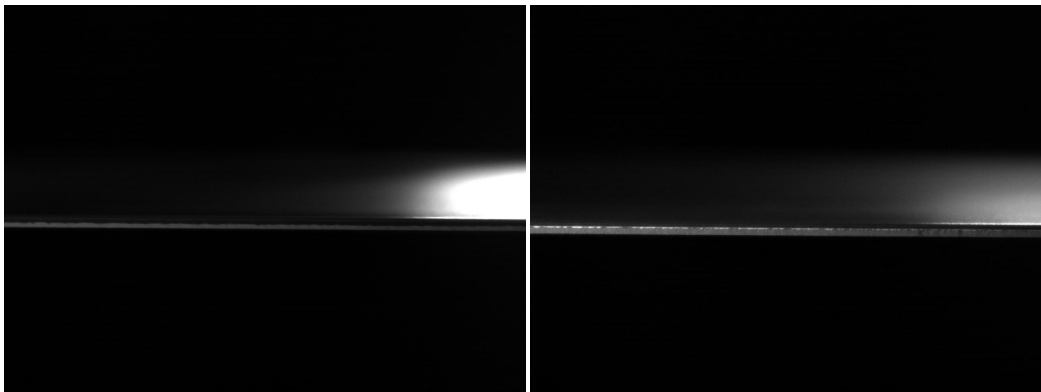
Obrázok 2 zobrazuje snímku hrany, ktorá je prekrytá červeným filtrom v oblasti, kde sa nachádzajú nedôležité pixely, zatiaľ čo oblasť obrázka prekrytá zeleným filtrom pokrýva dôležité pixely.



Obr. 2 Pomer dôležitých a nedôležitých pixelov v snímkach.

Obrázky 3 a 4 sú príklady dobrého a zlého orezu. Rozdiel medzi obrázkami je na prvý pohľad nebadateľný. Na snímke zlého orezu je možné vidieť malý strieborný pás na hrane orezu. Tento pás je v skutočnosti pílkovitosť plechu a strieborná farba je len odrazom svetla na ostrej hrane nekvalitne orezaného plechu.

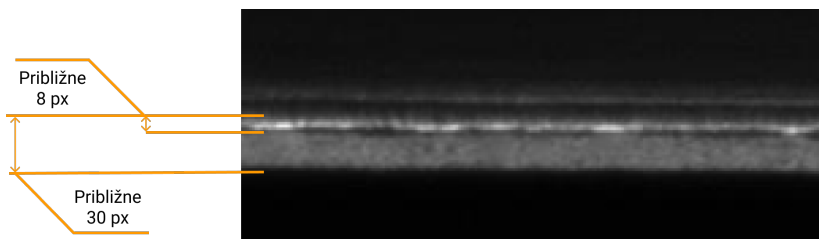
Na obrázku 5 môžeme vidieť hranu oceleového plechu získanú priblížením/zväčšením, na ktorom sa nachádza konštrukčná chyba. Túto konštrukčnú chybu nazývame pílkovitosť plechu a ide o



Obr. 3 Dobrý orez.

Obr. 4 Zlý orez.

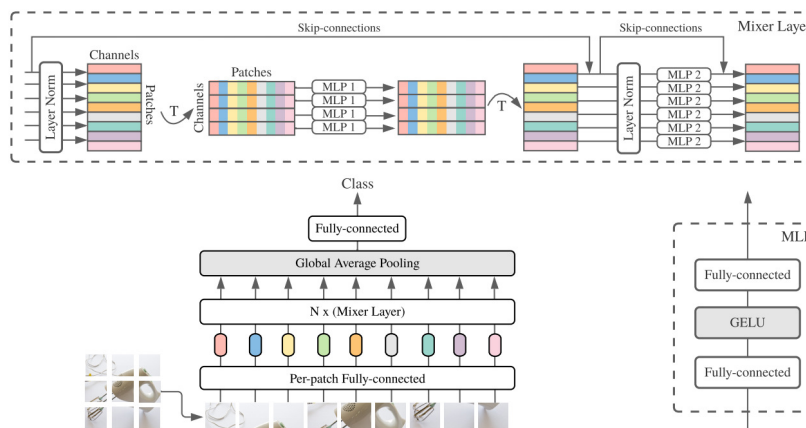
chybu ktorú sa budeme snažiť našim riešením včas odhaľovať. Na obrázku si môžeme všimnúť svetlú plochu od ktorej sa odráža svetlo. Pracovníci z moriacej linky túto plochu nazývajú aj strieborná niť, ihla alebo pílka. Práve táto svetlá časť je reprezentatívna časť obrázka prostredníctvom ktorej sme schopní klasifikovať orez ako dobrý alebo zlý. Ako môžeme vidieť na obrázku, reprezentatívna časť je zhruba 8 pixelov vysoká, čo znamená, že iba asi 1% snímky predstavuje časť, ktorá rozhoduje o kvalite orezu.



Obr. 5 Reprezentatívna časť obrázka orezu s vyznačením rozmerov (px - pixel).

B. Architektúra MLP mixer

Výskumníci z tímov Google Research a Google Brain Team prišli s článkom [7], v ktorom predstavili novú architektúru s názvom MLP mixer (Obr. 6). Táto architektúra nevyužíva konvolučné vrstvy, no napriek tomu dosahuje podobný výkon ako architektúra Vision Transformer (ViT) rozobraná v článku [8]. Navyše architektúra MLP mixer nevyužitím konvolučných vrstiev [9] ušetrila potrebný výkon pri tréňovaní a dokáže sa učiť trojnásobne rýchlejšie ako spomínaná architektúra ViT.



Obr. 6 Architektúra MLP mixer Zdroj: [7]

Architektúra MLP mixér sa skladá z troch hlavných častí. V prvej časti sa obrázok rozdelí na menšie časti tzv. kúsky (z angl. patch). Každý kúsok obrázku je vstupom do jednoduchej doprednej neurónovej siete, ktorej výstupom sú lokálne obrazové príznaky daného kúska. V druhej časti siete sú tieto lokálne príznaky spájané pomocou mixovacích vrstiev (z angl. mixer layer) do globálnych príznakov. V poslednej časti siete sa tieto globálne príznaky spolu s lokálnymi príznakmi (skrz skip prepojenia) použijú ako vstup do finálnej doprednej neurónovej siete ktorá vykonáva finálnu klasifikáciu.

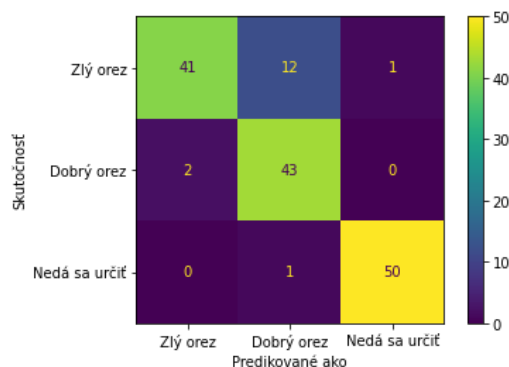
C. Vyhodnotenie klasifikačného modelu

Po niekoľkých iteráciách tréovania modelov kedy sme hľadali vhodné nastavenie veľkosti výrezov, hĺbok, a dimenzií, sme dospeli k modelu MLP mixér s nasledujúcimi vlastnosťami a hyper-parametrami, ktoré môžeme vidieť na obrázku 7.

Klasifikačný model verzia 2.1		
MLPMixer		
Vstup (1292px * 952px) Farebný model (Úroveň šedej)	dimension: 5 depth: 3 patch size: (64px * 64px) Počet tréovaných parametrov: 1.7 Milióna	Výstup Klasifikačné triedy: Zlý orez, Dobry orez, Nedá sa určiť
Veľkosť finálneho modelu: 20MB		

Obr. 7 Klasifikačný model MLP mixér

Modelom verzie 2.1, ktorého štruktúru vizualizuje obrázok 7, sme na testovacej sade dát dosahovali výsledky uvedené v tabuľke 1.



Tabuľka 1

Matica zámen klasifikačného modelu.

Na matici zámen môžeme vidieť väčšiu chybovosť predikcie iba pre triedu zlý orez, ktorý model označuje ako dobrý orez.

	presnosť	zapamätanie	f1-skóre	počet snímiek
Zlý orez	0.95	0.76	0.85	54
Dobry orez	0.77	0.96	0.76	45
Nedá sa určiť	0.98	0.98	0.98	51
celková presnosť			0.89	150
makro priemer	0.9	0.9	0.89	150
vážený priemer	0.91	0.89	0.89	150

Tabuľka 2

Výsledok meraní klasifikačného modelu

IV. AGREGAČNÁ FUNKCIA

O určenie kvality oceľovej hrany zvitku a aktivovanie alarmu v prípade zlej kvality sa stará agregáčnā funkcia. Jej algoritmus je spúšťaný periodicky v nastavenom čase T , a vykonáva sa súčasne pre ľavú a pravú orez. Pre aktivovanie alarmu je postačujúce, ak je znehodnotenā len jedna hrana oceľového pásu.

Parametre algoritmu ako je počet sledovaných snímkov N , prahovā hodnota výskytu chybného orezu H pre akosť G , alebo doba za ktorú sa spustí nasledujúca agregácia T , môžeme považovať ako premenné pre nastavenie agregáčného algoritmu.

Hodnoty týchto premenných získavame na základe zozbieraných dát o zvitkoch v našej dátovej sade. Ako sme spomínali, v aplikácií tabuľkového prehľadu operátori moriacej linky pravidelne označujú zvitky. Pri tomto zbere dát do našej dátovej sady zaznamenávame všetky dostupné štatistické údaje o zvitku, ako je percentuálny výskyt jednotlivých klasifikačných tried pre ľavú a pravú stranu. Taktiež záznamy o akosti, hrúbke, či šírke oceľového pásu. K týmto dátam operátori pridávajú označenie či pre takýto zvitok by chceli mať v budúcnosti aktivovaný alarm. Taktiež musia svoje rozhodnutie zdôvodniť, nakoľko podľa rozhodnutí vieme vytvárať nové triedy rozhodnutí, ktoré urýchľujú následný zber a zároveň ho robia prehľadnejším.

Motiváciou pre operátora moriacej linky označovať zvitky v našej aplikácii je skutočnosť, že čím viac dát o zvitkoch máme, tým sme schopní lepšie nastaviť agregáčnú funkciu. S každým dobrým označením zvitku operátor výroby napomáha k minimalizovaniu falošne pozitívnych spustení alarmu.

Algoritmus agregáčnej funkcie je nasledovný.

- 1) Stiahnutie všetky nových snímkov orezu od doby poslednej agregácie.
- 2) Iterácia nových snímkov a ich klasifikácia.
- 3) Priradenie klasifikovaných snímkov do zásobníka agregáčnej funkcie
- 4) Pozorovanie posledných N snímkov zo zásobníka. Na týchto snímkach sa kontroluje percentuálny počet vyskytnutých klasifikácií zlého orezu. Ak tento počet prekračuje nami nastavenú prahovú hodnotu H pre akosť G , zvitok sa ukladá do databázy so záznamom aktivovaného alarmu.

V. WEBOVÁ APLIKÁCIA

Samotná implementácia riešenia predstavuje webovú aplikáciu vhodnú na nasadenie v danom produkčnom prostredí. Webovā aplikácia pozostáva z niekoľkých aplikácií a slúži pre prehľadanie historických dát, zobrazenie živého záznamu kamier moriacich liniek (s prípadným alarmom) a taktiež pre označovanie a pridávanie nových dát do dátových sád.

A. Aplikáčnā časť 1: Živé zobrazenie

Tāto aplikácia je určenā pre operátorov moriacich liniek. Aplikácia zobrazuje aktuálne snímky oceľovej hrany, ktorá je v danom momente orezávanā na moriacej linke operátora. Okrem samotnej snímky sa operátorovi zobrazuje aj klasifikácia pre danú snímku. V prípade že agregáčnā funkcia vyhodnotí, že na danej moriacej linke došlo k zhoršeniu kvality orezu oceľového pásu, aplikácia spustí alarm. Alarm pozostáva zo zvukového a vizuálneho signálu. Pod vizuálnym signálom môžeme rozumieť pozadie aplikácie pulzujúce červenou farbou.

B. Aplikáčnā časť 2: Tabuľkový prehľad

Prostredníctvom aplikácie tabuľkový prehľad je používateľ schopný vyhľadať oceľový zvitok, ktorý prešiel procesom orezu na moriacej linke. Používateľ po vyhľadaní oceľového zvitku sa môže o zvitku dozvedieť informácie o jeho akosti, dátumu výroby, výstupu agregáčnej funkcie pre aktivovanie alarmu. Taktiež je možné prezerať všetky snímky orezu oceľového zvitku a taktiež ich klasifikáciu.

V prípade zlej klasifikácie snímku oceľovej hrany, je používateľ schopný tento snímok správne označiť a pridať do dátovej sady. Zároveň je používateľ schopný pridať zvitok do dátovej sady s označením aktivovania alebo neaktivovania alarmu.

C. Aplikáčnā časť 3: Prehrávač

Pomocou aplikácie prehrávač je používateľ schopný zvoliť jeden z oceľových zvitkov, a následne spustiť prehrávanie pre simuláciu živého zobrazovania. Pomocou simulácie živého zobrazenia môžeme testovať klasifikačný model a agregáčnú funkciu na historických dátach a reprodukovať prípadné zlyhanie klasifikácie.

VI. ZÁVER

Cieľom článku bolo prezentovať proces návrhu systému, ktorý by bol schopný včas upozorniť operátora výroby na moriacej linke na zhoršenú kvalitu orezu moriacej linky, a tak dosiahnuť ekonomickú a zároveň ekologickú úsporu miernou formou modernizácie. Naša implementácia neohrozuje pracovné miesto operátora na danej linke – napomáha však zvýšeniu efektívnosti jej/jeho práce.

Výsledkom našej práce je klasifikačný model, ktorý je schopný na základe snímky z kamery určiť kvalitu orezu ocelového pásu do troch kategórií kvality. Výstupom je taktiež agregáčna funkcia ktorá sleduje klasifikačné skóre modelu, a jej výstupom je alarmový signál. Tieto dve časti spája webová aplikácia, ktorá ponúka operátorovi moriacej linky živý prehľad aktuálneho orezu s klasifikáciou orezu a vizuálnym a zvukovým signálom alarmu.

Riešenie ostáva otvorené voči vylepšeniam klasifikácie napr. občasným pretrénovaním na základe doplnených označení v sade dát. Ako väčšina podobných aplikácií, aj to naše vykazuje silnú závislosť na rozlíšení snímok, preto vnímame spracovanie vstupov systému ako najviac náchylné k zmenám v prípade zásahu do prostredia nasadenia.

Naša aplikácia bola nasadená do prevádzky moriacich liniek, v spoločnosti U.S.Steel Košice a.s., a to integráciou s existujúcim vybavením snímania resp. zobrazenia snímok. Nasadenie aplikácie neovplyvní pracovnú náplň zamestnanca, ktorý ju používa. Až dlhší čas prevádzky môže odhaliť mieru dosiahnutých úspor.

POĎAKOVANIE

Autori ďakujú členom tímu Digital studio zo spoločnosti U.S.Steel Košice a.s. za odbornú konzultáciu a za poskytnutie údajov potrebných pre vypracovanie praktickej implementácie.

LITERATÚRA

- [1] Lýdia Sobotová a Ľudmila Dulebová, "Vplyv opotrebenia pracovných valcov na kvalitu morených ocelových plechov," *TRANSFER INOVÁCIÍ*, č. 13, 2009.
- [2] A. Jayal, F. Badurdeen, O. Dillon, a I. Jawahir, "Sustainable manufacturing: Modeling and optimization challenges at the product, process and system levels," *CIRP Journal of Manufacturing Science and Technology*, zv. 2, č. 3, st. 144–152, 2010, sustainable Development of Manufacturing Systems. [Online]. Dostupné na odkaze: <https://www.sciencedirect.com/science/article/pii/S1755581710000131>
- [3] A. Lebar, M. Junkar, A. Poredoš, a M. Cvjeticanin, "Method for online quality monitoring of awj cutting by infrared thermography," *CIRP Journal of Manufacturing Science and Technology*, zv. 2, č. 3, st. 170–175, 2010, sustainable Development of Manufacturing Systems. [Online]. Dostupné na odkaze: <https://www.sciencedirect.com/science/article/pii/S1755581710000118>
- [4] Salman Khan a kol., *A Guide to Convolutional Neural Networks for Computer Vision*. Morgan and Claypool, 2018.
- [5] Elena Šikudová a kol., *Počítačové videnie. Detekcia a rozpoznávanie objektov*. Wikina, 2013.
- [6] Abhiram Natarajan a kol., "An approach to real time parking management using computer vision," in *Proceedings of the 2nd International Conference on Control and Computer Vision*, sér. ICCCV 2019. New York, NY, USA: Association for Computing Machinery, 2019, s. 18–22. [Online]. Dostupné na odkaze: <https://doi.org/10.1145/3341016.3341025>
- [7] Ilya O. Tolstikhin a kol., "MLP-Mixer: An all-MLP Architecture for Vision," *CoRR*, zv. abs/2105.01601, 2021. [Online]. Dostupné na odkaze: <https://arxiv.org/abs/2105.01601>
- [8] Alexey Dosovitskiy a kol., "An image is worth 16x16 words: Transformers for image recognition at scale," *CoRR*, zv. abs/2010.11929, 2020. [Online]. Dostupné na odkaze: <https://arxiv.org/abs/2010.11929>
- [9] Halima Bouzidi, Hamza Ouarnoughi, Smail Niar a Abdessamad Ait El Cadi, "Performances Modeling of Computer Vision-Based CNN on Edge GPUs," *ACM Trans. Embed. Comput. Syst.*, Marec 2022. [Online]. Dostupné na odkaze: <https://doi.org/10.1145/3527169>

Návrh mobilnej aplikácie pre získavanie dát z monitorovacej stanice

¹Samuel ANDREJČÍK, ²Luboš OVSEŇÍK, ³Jakub ORAVEC, ⁴Maroš LAPČÁK,
⁵Norbert ZDRAVECKÝ

^{1,2,3,4,5} Katedra elektroniky a multimediálnych telekomunikácií, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

¹samuel.andrejck@tuke.sk, ²lubos.ovsenik@tuke.sk, ³jakub.oravec@tuke.sk,
⁴maros.lapcak@tuke.sk, ⁵norbert.zdravecky@tuke.sk

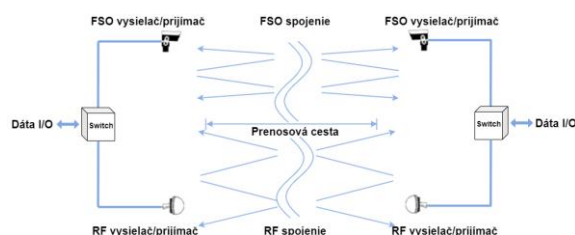
Abstrakt — Predmetom tohto príspevku je návrh mobilnej aplikácie určenej pre analýzu dát získaných prostredníctvom senzorov monitorujúcich poveternostné podmienky počasia. Tieto dáta sú potrebné k správne predikovaní prepínania liniek v hybridnom FSO/RF prenosovom systéme.

Kľúčové slová — FSO/RF, mobilná aplikácia, optická komunikácia voľným prostredím, Unity Engine, Unity 3D

I. ÚVOD

V roku 2008 bola spustená prvá verzia obchodu App store, čo spustilo revolúciu v oblasti mobilných zariadení. Okrem bežných, výrobcami predinštalovaných aplikácií sme tak dostali stovky nových možností, napr. rôzne hry, aplikácie obchodov uľahčujúce online nákupy, správy a nástroje, ako kalkulačky počítajúce kalórie či dokonca kompas. Všetko čo sme „potrebovali“ sa nachádzalo v našom mobilnom zariadení. Nie len že sa tento trend nezastavil, ale raketovo napreduje ďalej. V súčasnosti používame aplikácie na báze strojového a hĺbkového učenia v podobe osobných asistentov, ktorí s nami dokážu viesť rozumný dialóg. Informácie ako také sa stali najdôležitejšou súčasťou nášho života. Aplikácie nám slúžia ako prehliadače týchto informácií, čo nám uľahčuje každodenný život. Vďaka spravodajským aplikáciám vieme, čo sa deje vo svete, aplikácie internetového bankovníctva nám ukazujú stav účtu a meteorologické aplikácie nám pomáhajú s výberom primeraného odevu.

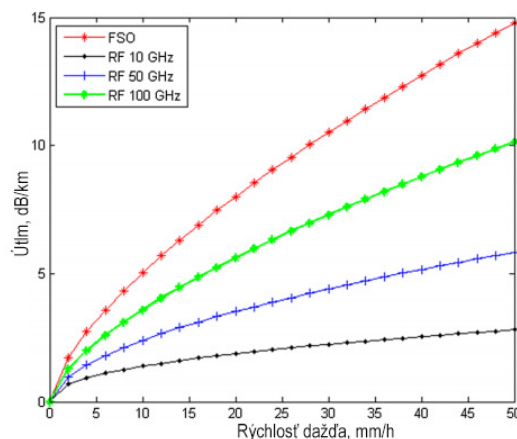
Predstavovaná aplikácia pre analýzu dát získaných prostredníctvom senzorov monitorujúcich poveternostné podmienky počasia slúži k správne nastaveniu vhodnej prenosovej cesty v hybridnom FSO/RF prenosovom systéme. FSO (Free Space System) je systém pre optický prenos dát voľným prostredím (vzduchom), ktorý funguje na podobnom princípe ako prenos dát optickými vláknami a teda tak, že informácie sú prenášané pomocou lúčov svetla. Hlavným rozdielom medzi týmito systémami je prostredie, v ktorom je tento prenos realizovaný. Zatiaľ čo optické vláknové spoje na prenos používajú najčastejšie optické vlákna na báze SiO₂, ktorých nevýhodou je ich cena a náročná inštalácia, FSO používa atmosférický nelicencovaný kanál, vďaka čomu je tento systém lacnejší a ľahko inštalovateľný pri porovnateľných prenosových rýchlostiach a šírke prenosového pásma. Nevýhodou je však nutnosť priamej viditeľnosti medzi FSO hlavicami a pomerne krátkej vzdialenosti (cca 20km), na ktorej operujú. Vplyv počasia brániaci priamej viditeľnosti tak môže mať negatívne účinky. Doplnením systému prenosu dát voľným prostredím o prenos dát rádiovým RF signálom (Radio Frequency) vzniká hybridný FSO/RF systém (Obr. 1), vyzdvihujúci dobré vlastnosti jednotlivých systémov, čím sa stáva odolnejším voči vplyvom počasia [1].



Obr. 1 Hybridný FSO/RF systém.

II. HYBRIDNÉ FSO/RF PRENOSOVÉ SYSTÉMY

FSO/RF je spojenie dvoch prenosových systémov, určených na prenos dát voľným prostredím využívajúcich nelicencované frekvenčné pásma. Hybridná FSO/RF komunikácia je implementovaná tak, aby kombinovala výhody oboch systémov pre zabezpečenie spoľahlivého rýchleho širokopásmového komunikačného spojenia. Aby sme neznehodnotili výhody FSO systému, je potrebné nájsť kompromis v nastaveniach hybridnej siete. RF linka by mala byť dostupná hneď v momente, ako sa FSO stáva nedostupnou. Šírka pásma RF linky by mala dosahovať približné hodnoty FSO linky [1]. Veľké rozdiely v šírke pásma by spôsobovali zníženie prenosovej rýchlosti pri prepnutí z FSO na RF linku. Neexistuje však „najlepšie“ nastavenie hybridného prenosového systému, všetko to závisí od konkrétnej situácie a požiadaviek, ako aj konkrétneho miesta implementácie. Napríklad kombinácia FSO/RF s vysokými prenosovými rýchlosťami, ktoré si vyžadujú veľké nosné frekvencie nie sú odporúčané v kontinentálnom podnebí, pretože sú obe ovplyvňované hmlou, dažďom, či snehom. Naopak, implementáciou tejto kombinácie v tropickom suchom podnebnom pásme môžeme získať stabilnú komunikačnú sieť s vysokými prenosovými rýchlosťami. Komerčné RF linky sa pohybujú vo frekvenčných pásmach od 60 GHz do 80 GHz. Frekvencie okolo 60 GHz sú nelicencované, no zároveň nemôžu byť použité na väčšie vzdialenosti kvôli vysokému útlmu ovzduším, konkrétne kyslíkom, ktorý spôsobuje absorpciu až 15 dB/km. Frekvencie od 70 GHz až 80 GHz RF linky používané modernejšími zariadeniami dokážu prenášať dáta na vzdialenosti do 3 km rýchlosťou 1Gb/s nezávisle od počasia, avšak tieto frekvencie bývajú dostupné až po získaní licencie. Použitím nižších frekvencií, ktoré nevyžadujú licencie sa znižujú aj prenosové rýchlosti, no výhodou je malý vplyv atmosférických javov na prenos. Na Obr. 2 môžeme vidieť graf útlmu vplyvom dažďa na jednotlivé prenosové systémy za použitia nastavení z Tabuľky 1.



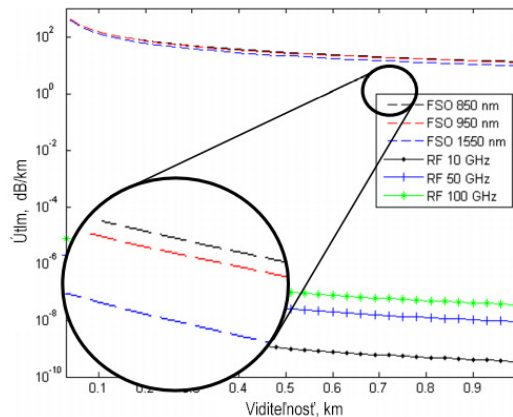
Obr. 2 Graf útlmu vplyvom dažďa.

Tabuľka 1

(a) Parametre FSO systému, (b) Parametre RF systému

Parameter	Hodnota	Parameter	Hodnota
Výkon vysielača	5 mW	Výkon vysielača	5 mW
Uhol divergencie laserového lúča	1 mrad	Účinnosť vysielača	0,5
Účinnosť vysielača	0,9	Účinnosť prijímača	0,7
Účinnosť prijímača	0,9	Vlnová dĺžka	3 cm (100 GHz)
Vlnová dĺžka	850, 950, 1500nm	Priemer antény vysielača	0,18 m
Dĺžka vedenia (L)	0,1 km ≤ L ≤ 1 km	Priemer antény prijímača	5 m
Viditeľnosť	1 km	Citlivosť prijímača	-110 dBm
Priemerná atmosférická teplota	20 °C	Ekvivalentný odpor	50 kΩ
Priemerná relatívna vlhkosť vzduchu	67,9 %	Teplota systému	290 K
Úhrn zrážok	18,3 mm/h		
Citlivosť prijímača	-30 dBm		
Rýchlosť prenosu dát	100 Mbps		
Priemer prijímača	15 cm		
Absolútna teplota fotodiódy	298 K		
Šírka pásma	0,5 GHz		

Zvyšujúcou sa intenzitou dažďových zrážok sa zvyšuje aj útlm oboch prenosov v rôznych frekvenciách. Ako sme už spomínali, RF prenos s frekvenciou nižšou ako 20 GHz vykazuje nižší útlm, ako pri frekvenciách vyšších ako 20 GHz. Na Obr. 3 môžeme vidieť útlm podľa viditeľnosti spôsobeného hmlou. Väčšia vlnová dĺžka FSO systému znižuje negatívny dopad hmly, no zďaleka nedosahuje hodnoty RF systému pri frekvencii 10 GHz. Za nepriaznivých podmienok je FSO/RF hybridný systém prepnutý na záložnú linku, ktorou je práve RF linka. Vďaka tomu dokážeme udržať komunikáciu aj za zlého počasia [2].



Obr. 3 Graf útlmu vplyvom hmly.

III. NÁVRH MOBILNEJ APLIKÁCIE PRE MONITOROVANIE DÁT

Pre tvorbu aplikácie monitorujúcej počasie sme používali program pre tvorbu hier a aplikácií Unity Engine [4, 5], Photoshop [6] pre tvorbu grafiky a Visual Studio [3] pre písanie skriptov v programovacom jazyku C#. Samotnú tvorbu aplikácie by sme si mohli rozdeliť na niekoľko častí:

- získanie dát, roztriedenie a ich následné uloženie,
- tvorba grafov a zobrazenie údajov,
- užívateľské rozhranie.

A. Získanie dát, roztriedenie a ich následné uloženie

Dáta zo senzorov umiestnených na jednodoskovom počítači Raspberry pi sú ukladané do databázy, z ktorej ich budeme sťahovať pomocou http požiadavky. Stiahnutý súbor s príponou .csv obsahuje údaje o dátume a čase vo formáte mm/dd/yyyy hh:mm, teplote meraní z viacerých senzorov, tlaku, vlhkosti, rýchlosti vetra a prijatom optickom výkone. Na Obr. 4 môžeme vidieť príklad jedného merania senzormi, uloženého v .csv súbore.

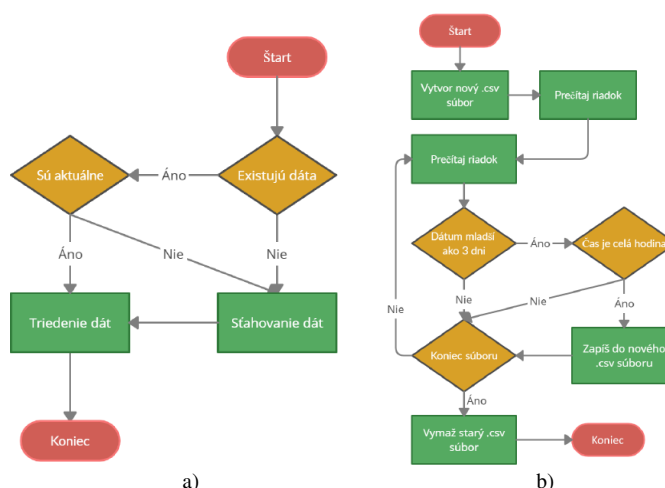
```
9/21/2016 7:51,-39.66666667,17.4,990.405,15.4,49.099998,11.125,0.488776,0.399637,12,4000
```

Obr. 4 Príklad sťahovaných dát.

Po zapnutí aplikácie sa spustí skript, ktorý má za úlohu zistiť dostupnosť dát. Ak sa v mobilnom zariadení nenachádza .csv súbor, je nútený si ho stiahnuť z databázy. Tú vieme pohodlne meniť v nastaveniach aplikácie. Ak už dáta existujú, aplikácia si overí ich aktuálnosť porovnaním času zariadenia s časom vytvorenia súboru, v ktorom sa dáta nachádzajú. Pri ich zhode vieme, že sú dáta aktuálne a môžeme prejsť k ich triedeniu. Ak sa časy nerovnajú, znamená to, že naše dáta sú už viac ako hodinu staré a tak ich znova stiahneme. Tento skript sa spúšťa pri každom spustení scény „Aktuálne počasie“, to znamená pri zapnutí, ako aj pri prepnutí z iných scén aplikácie. Pri opätovnom sťahovaní sa najprv vymažú staré dáta, aby nedochádzalo k ich zbytočnému zhromažďovaniu a zahlcovaniu pamäte a až potom sa stiahnu nové.

Triedenie dát po ich stiahnutí sa vykonáva v tom istom skripte, súbor je postupne prechádzaný zvrchu nadol. Prvý riadok .csv súboru obsahujúci názvy hodnôt je preskočený, nasledujúce riadky s hodnotami sú čítané a rozanalyzované. Prečítaný riadok je rozdelený na jednotlivé hodnoty v dátovom type „string“, ktoré sú oddelené čiarkou. Prvým údajom v riadku je dátum a čas vo formáte dd/mm/yyyy hh:mm. Následne je tento reťazec prevedený do dátového typu „date“. Podľa dátumu a času určujeme čo so všetkými hodnotami v riadku urobíme. Ak je čas v analyzovanom riadku celá hodina, teda napríklad 8:00, 9:00 a podobne je tento riadok ďalej spracovávaný, v inom prípade ho zahodíme a prejdeme na ďalší. Pre túto aplikáciu sme sa rozhodli zobrazovať dáta pre tri dni - aktuálneho a z dvoch predošlých dní a to 24 hodnôt za deň. V konečnom dôsledku dostaneme grafy s teplotou, tlakom, vlhkosťou vzduchu a prijatým

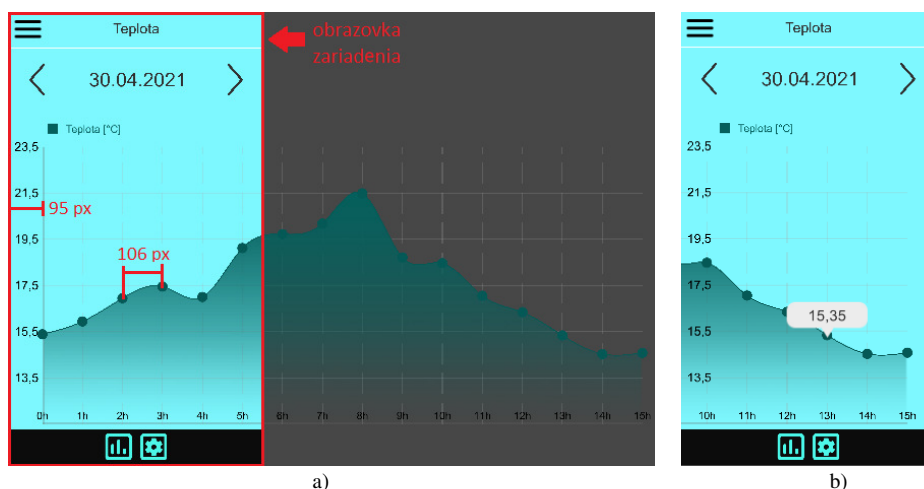
optickým výkonom v hodinových intervaloch. Tým, že je sťahovanie aj čítanie dát vykonávané pri každom načítaní scény, bolo by veľmi neefektívne prechádzať celý súbor obsahujúci niekoľko tisíc riadkov údajov. Napríklad súbor dát z 20 dní by obsahoval $20 \text{ dní} * 24 \text{ hodín} * 60 \text{ minút}$, čo sa rovná 28800 riadkov, za predpokladu, že merania sú opakované každú minútu. Celý tento proces by zaberol zbytočne veľa zdrojov a v konečnom dôsledku by spôsoboval spomalenie systému, sekacie, či pád aplikácie. Aby sme predišli tomuto scenáru, je pri triedení dát vytvorený úplne nový .csv súbor, data.csv. Ak sa už v zdrojoch aplikácie nachádza súbor s týmto menom, je zmazaný a nahradený novým. Ďalej pri triedení sú doň prepisované vyhovujúce dáta zo súboru stiahnutého zo servera. Nami zvolené trojdňové zobrazenie údajov o počasi bude obsahovať nie viac ako $3 \text{ dni} * 24 \text{ hodín}$, teda 72 riadkov. Po ukončení triedenia dát sa súbor stiahnutý zo servera odstráni a nový .csv súbor uloží. Týmto nepredídeme len zbytočnému čerpaniu zdrojov, ale dokážeme aj opätovne rýchlo načítať a efektívne ukladať dáta. Celý tento algoritmus je zavŕšený pridelením času vytvorenia k novému .csv súboru. Ten slúži na zistenie aktuálnosti dát, ako je už opísané a naznačené na Obr. 5.



Obr. 5 (a) Rozloženie grafu, (b) Graf scény „Teplota“.

B. Tvorba grafov a zobrazenie údajov

Keďže vytvárame aplikáciu pre mobilné zariadenia musíme dbať na jej prehľadnosť. Preto sme sa rozhodli zobrazovať údaje v hodinových intervaloch a veľkosť grafu prispôbovať k počtu hodnôt na x-ovej osi. Avšak pri plnom grafe, teda 24 hodnôt by sa nám aj tak celý nezместil na obrazovku mobilného telefónu. To sme vyriešili jeho „scrollovaním“. Hodnoty sú do grafu ukladané postupne podľa toho, ako sú čítané zo súboru. Každá z nich je oddelená od predchádzajúcej 106 pixelov a o presne tú istú hodnotu je inkrementovaná dĺžka grafu (Obr. 4a).



Obr. 4 (a) Rozloženie grafu, (b) Graf scény „Teplota“.

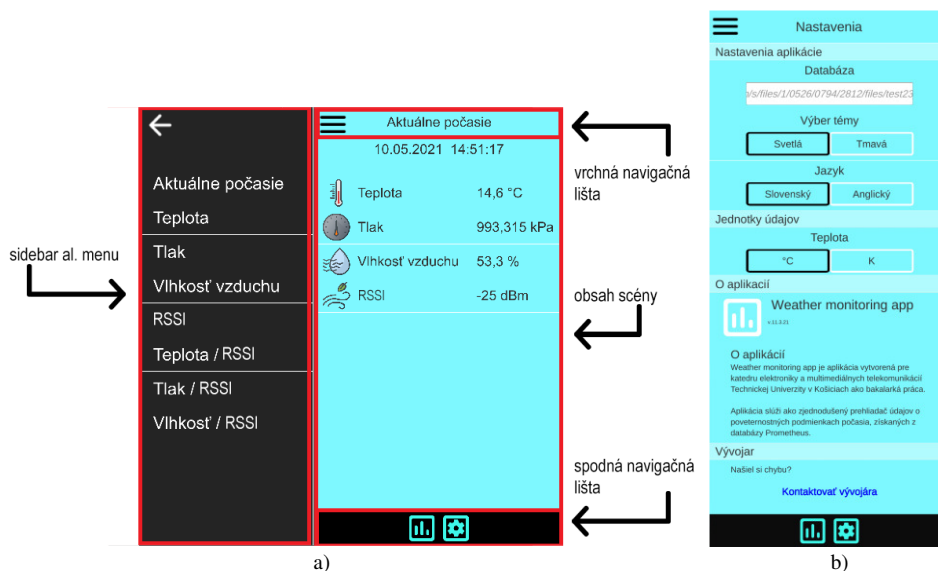
Ak jeho dĺžka prekročí šírku displeja, graf sa stáva „scrollovatelným“ a zachádza až za zorné pole zariadenia. Jednoduchým posúvaním grafu vpravo, či vľavo si vieme zobraziť namerané hodnoty pre daný čas. Aby sme predišli neustálemu posúvaniu a hľadaniu aktuálneho času a k

nemu prislúchajúcej hodnoty, graf sa pri zapnutí scény automaticky posunie na poslednú hodnotu, ktorá je zároveň najaktuálnejšou. To sme dosiahli vynásobením počtu hodnôt číslom 106 pixelov, čo je vzdialenosť medzi nimi a pripočítali 95 pixelov, to je nemenná vzdialenosť medzi ľavou stranou obrazovky a začiatkom grafu (Obr. 4a). Samozrejmosťou je zafixovanie y-novej osi na ľavej strane obrazovky, vďaka čomu vieme približne určiť hodnoty jednotlivých bodov. Pre zistenie presnej hodnoty je potrebné stlačiť na bod kedy nám vyskočí správa s presným číselným údajom, ako to môžeme vidieť na Obr. 4b. Kvôli zachovaniu jednoduchosti a prehľadnosti sú čísla v grafoch zobrazené bez jednotiek, tie je vidieť v legende vľavo hore. V našej aplikácii budeme zobrazovať grafy pre teplotu, tlak, vlhkosť vzduchu, prijatý optický výkon a ďalšie.

C. Užívateľské rozhranie

Celá aplikácia pozostáva zo scén, tie by sme si mohli predstaviť ako jednotlivé stránky webu, ktoré sú samostatným funkčným pilierom, no ich spojenie vytvára komplexnú web stránku. Každá zo scén je takýmto pilierom skladajúceho aplikáciu. Scény sa skladajú z objektov a tie v konečnom dôsledku tvoria užívateľské rozhranie. V našej aplikácii rozlišujeme tri druhy scén: a to *aktuálne počasie*, *nastavenia* a *grafy*.

Myšlienka scény **aktuálneho počasia** (Obr. 5a) je poskytnúť používateľom informácie čo najrýchlejšie a to hneď po zapnutí aplikácie. Na tejto stránke sa zobrazujú posledné namerané hodnoty spolu s aktuálnym časom a dátumom. Tento časový údaj je len orientačný a závisí na systémovom nastavení zariadenia. Pod ním sa nachádzajú hodnoty teploty, tlaku, vlhkosti vzduchu a veľkosti prijatého optického výkonu. Pri každej z nich je malá ikona predstavujúca danú jednotku, ikony sme stiahli z obchodu Unity a následne upravili v programe pre úpravu obrázkov, Photoshop. Táto scéna taktiež obsahuje skryté objekty, ktoré je možné vidieť až po istej interakcii. Ide o textové pole hlásiace chyby pri čítaní, alebo sťahovaní dát. Nachádza sa priamo pod hodnotami a zobrazí sa práve vtedy, keď sa nám nepodarí stiahnuť súbor dát z databázy. Upozorní nás na to správou: „Nastala chyba pri sťahovaní dát, prosím skontrolujte svoje internetové pripojenie.“, ak sa nám však podarí stiahnuť súbor, ale nedokážeme z neho získať žiadne údaje, dostaneme správu s textom: „Pre dnešný deň neexistujú žiadne dáta.“. Tento prípad nastane práve vtedy, ak sme súbor dokázali stiahnuť, prečítať, no nenachádzajú sa v ňom dáta pre aktuálny deň.

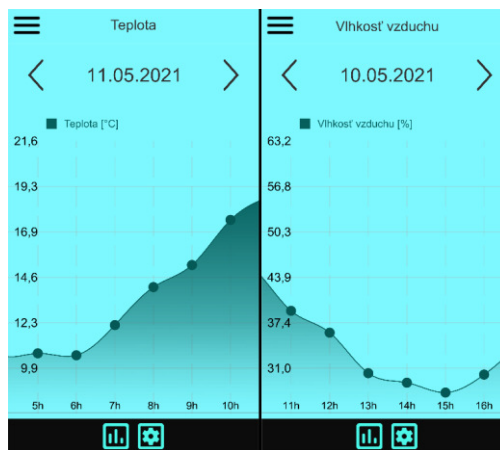


Obr. 5 Scéna (a) „Aktuálneho počasia“, (b) „Nastavenia“.

Scéna **nastavení** je ďalším dôležitým prvkom aplikácie. Tu si dokážeme nastaviť jednotky zobrazovaných hodnôt, jazyk, nachádzajú sa tu tiež tlačidlá pre výber témy a textové pole slúžiace na vloženie linku odkazujúceho na databázu. Spolu s týmito nastaveniami sa tu nachádzajú informácie o aplikácii ako aj kontaktného formulára v prípade nájdenia chyby v aplikácii. Ako môžeme vidieť na Obr. 5b, všetky tieto nastavenia zaberajú veľa miesta a tak sa nezmestia na obrazovku s veľkosťou pod 5,2 palcov vrátane. To sme vyriešili podobne ako pri grafoch s hodnotami presahujúcimi obrazovku, teda ak nastavenia presahujú rozsah obrazovky, obsah tejto scény sa stáva posúvateľným. Súvisiacim prvkom s rôznymi veľkosťami displeja je jeho orientácia. Tá je predvolene nastavená aby sledovala gyroskop mobilného zariadenia a podľa

neho otáčala aplikáciu, vždy tak dostaneme scénu s vrchnou navigačnou lištou smerujúcou hore, nezávisle od vertikálneho otočenia smartfónu. Toto nastavenie však spôsobuje problémy s responzívnosťou, tá však funguje správne. Pri otočení displeja do horizontálnej polohy sa otočí aj všetok obsah vrátane vrchnej a spodnej navigačnej lišty, tie sa natiahnu na šírku celej obrazovky pričom pre obsah ostane široké a nízke miesto. Grafy sa tomuto obsahu prispôbia a dostávajú sa tak široké, nízke grafy so zlou čitateľnosťou. Pretože je orientácia nepodstatná v porovnaní s responzívnosťou nastavili sme ju na možnosť „portrait“, čo nám zabezpečí stálu orientáciu smerujúcu od prednej kamery smerom k mikrofónu.

Aj keď je graf najdôležitejšou súčasťou týchto scén a aplikácie celkovo, scény **grafov** (Obr. 6) obsahujú ďalšiu komponent. Kalendár, slúži na zobrazenie hodnôt z predošlých dní, čo znamená, že je to komponent priamo ovplyvňujúci graf. Jeho dátum, rovnako ako aj dátum v scéne aktuálne počasie závisí od systémového nastavenia telefónu. Pre správne zobrazenie dát je teda potrebné mať nastavené rovnaké časové pásmo, ako má stanica merajúca poveternostné podmienky počasia, GMT +2. Kalendárom si môžeme pozrieť počasie dva dni dozadu, spolu s aktuálnym dňom - aplikácia tak poskytuje dáta pre tri dni.



Obr. 6 Uživatelské rozhranie scén „Grafovy“.

IV. ZÁVER

Aplikáciu sa nám podarilo úspešne otestovať s využitím dát z meteorologickej stanice nachádzajúcej sa v areáli TUKE. Aplikáciu je možné používať ihneď a v prípade potreby je možný jej ďalší vývoj. Jednou z možností je pridanie scén s predpoveďou počasia a radarových snímok pre danú oblasť, čo by vylepšilo jej vlastnosti a okrem určenia príčin výpadku prenosu, by sme vedeli predpovedať ďalšie možné výkyvy a výpadky. Ďalšími vylepšeniami pre túto mobilnú aplikáciu by mohli byť takzvané push notifikácie zobrazujúce sa v stavovom riadku mobilného zariadenia, ktoré by upozorňovali na možné výpadky systému v závislosti od predpovedi počasia. Taktiež by aplikácia mohla byť doplnená o rôzne animácie počasia alebo animácie prechodu medzi scénami pre jemnejší a plynulejší vzhľad.

POĎAKOVANIE

Táto práca bola podporená výskumným projektom FEI-2022-84 "Techniky spracovania dát vo vysokorychlostných prenosových systémoch".

LITERATÚRA

- [1] M. Lapčák, et al., "Zaznamenávanie poveternostných podmienok prostredia pomocou monitorovacej stanice a parametra RSSI pre spôsob tvrdého prepínania v hybridných FSO/RF systémoch", 2020. <http://webcache.googleusercontent.com/search?q=cache:gGuEs9qSFRgJ:www.elektrorevue.cz/cz/download/zazna-menavanie-poveternostnych-podmienok-prostredia-pomocou-monitorovacej-stanice-a-parametra-rssi-pre-sposob-tvrdeho-prepinania-v-hybridnych-fso-rf-systemoc>.
- [2] N. Bonfiglio, dailydot.com, "DeepMind partners with gaming company for AI research", 2018. <https://www.dailydot.com/debug/unity-deempind-ai/>
- [3] A. Aggarwal, geeksforgeeks.org, "Introduction to Visual Studio", 2019. <https://www.geeksforgeeks.org/introduction-to-visual-studio/>
- [4] Technologies, Unity. unity.com. Unity Real-Time Development Platform | 3D, 2D VR & AR Engine. <https://unity.com/>
- [5] Technologies, Unity. unity3d.com. Quick guide to the Unity Asset Store. <https://unity3d.com/quick-guide-to-unity-asset-store>
- [6] Adobe. adobe.com. <https://www.adobe.com/sk/products/photoshop.html>

Evolúcia obrazu použitím evolučného algoritmu

¹Tomáš HUDÁK, ²Marián MACH

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

¹tomas.hudak@student.tuke.sk, ²marian.mach@tuke.sk

Abstrakt — Článok sa zameriava na problematiku využitia počítačov v oblasti vizuálneho umenia s dôrazom na algoritmickú tvorbu obrazov. Pozornosť je zameraná na aproximáciu cieľového obrazu pomocou navzájom sa prekrývajúcich geometrických primitív. Článok sa zaoberá použitím evolučného algoritmu pre vytváranie generovaných obrazov na základe trojuholníkových a obdĺžnikových primitív. V rámci experimentovania bolo realizovaných niekoľko experimentov s týmito geometrickými primitívami a spôsobom ich vzájomného skladania na vybranom cieľovom obraze.

Kľúčové slová — evolučný algoritmus, počítačové umenie, generovanie obrazu

I. ÚVOD

Jedna z oblastí využitia počítačov je oblasť počítačového umenia. To je definované ako „sluchový alebo vizuálny výtvor, ktorý bol vytvorený pomocou počítača“. Vo vizuálnej doméne sa jedná o pomerne širokú oblasť, kde na jednom konci spektra je použitie kresliaceho programu pre manuálne vytvorenie obrazu pomocou vhodnej periférie (myš, dotykový display, grafický tablet), zatiaľ čo opačný koniec spektra je reprezentovaný algoritmickým vytváraním obrazu (napríklad „Line Art“ tvorený iba čiarami alebo „ASCII Art“ formujúci obraz pomocou tlačiteľných znakov).

Jednou podoblasťou algoritmického vytvárania obrazov je evolučné vytváranie – obraz nevzniká naraz v jednom kroku, ale v mnohých krokoch. Je podrobený evolučnému vývoju, keď v jednom evolučnom kroku dochádza k nejakej jeho malej zmene. Akumuláciou veľkého počtu malých zmien sa postupne vyvíja z prvotného tvaru (často náhodného) k nejakému konečnému tvaru. Tento evolučný proces smeruje k cieľu, ktorým môže byť splnenie nejakých podmienok, dosiahnutie nejakej kvality alebo priblíženie sa k nejakému vzoru.

V prípade približovania sa k vzoru môže týmto vzorom byť cieľový obraz, ktorý je generovaným obrazom nejakým spôsobom aproximovaný. Príkladom je „TSP Art“ - nahradenie obrazu množinou bodov (ich hustota v nejakom segmente obrazu závisí od farby daného segmentu), ktoré sú prepojené čiarou tak, aby tá čiara bola čo najkratšia [1]. Iným prípadom je „Polygon Art“ - náhrada obrazu množinou jednofarebných segmentov (určeného tvaru a veľkosti), ktoré sa môžu rozličným spôsobom prekrývať a kombinovať. Príkladom je použitie algoritmu simulovaného žihania [2] či genetického algoritmu [3][4].

Článok sa zaoberá použitím evolučného algoritmu pre generovanie obrazu zloženého z geometrických primitív tak, aby sa síce výsledný obraz ponášal na zadaný cieľový obraz avšak nebol jeho presnou kópiou, skôr zmeneným spôsobom pohľadu.

II. GENEROVANIE OBRAZOV

Pre generovanie sme sa sústredili na prehľadávanie priestoru možných obrazov pomocou evolučného algoritmu, ktorý poskytuje efektívne riadený pohyb v tomto priestore [5]. Navrhli sme jednotlivé bloky algoritmu pre riešenie evolúcie obrazov [6].

A. Reprezentácia obrazov

Jednotliví členovia postupne sa vyvíjajúcej populácie reprezentujú generované obrazy. Samotný obraz je vytváraný ako RGB matica (bol použitý RGB farebný model), reprezentujúca výsledok vrstvenia obrazových segmentov tvorených geometrickými primitívami. V experimentoch sme použili dva typy týchto geometrických primitív – trojuholníkové

a obdĺžnikové primitíva. V zásade je možné každého jedinca reprezentovať dvomi spôsobmi:

- zoznamom použitých geometrických primitív,
- priamo RGB maticou.

Pri prvom spôsobe postačuje pre každé primitívum uchovávať súradnice (troch alebo štyroch) vrcholových bodov. Výhodou je jednoduchá realizácia genetických operátorov, umožňujúca širokú paletu možných zmien manipuláciou na úrovni jednotlivých primitív (napríklad zmena primitíva, pridanie alebo vypustenie primitíva, výmena primitív medzi dvomi obrazmi, ap.). Na druhej strane takúto reprezentáciu je potrebné v konečnom dôsledku transformovať na výslednú RGB maticu, čo pri veľkom počte primitív predstavuje značnú výpočtovú záťaž.

Pri druhom spôsobe nie je potrebná uvedená transformácia a je možné priamo pracovať s obrazom na úrovni pixelov (čo je potrebné pri porovnávaní cieľového a generovaného obrazu). Problémom je však práca genetických operátorov na úrovni jednotlivých primitív, ktoré nie sú reprezentované explicitným spôsobom.

B. Skladanie geometrických primitív

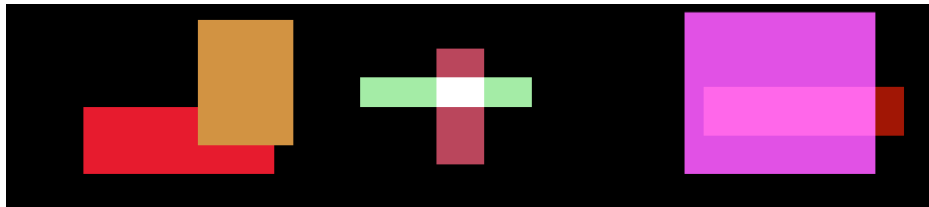
Pri vkladaní geometrického primitíva do obrazu sa prechádza každý pixel obrazu, ktorý je pokrytý plochou vkladaneho primitíva, pričom hodnota týchto pixelov je upravovaná. Každý takto upravovaný pixel má k dispozícii dve hodnoty:

- pôvodná hodnota pixelu (buď pozadie na ktorom je vytváraný obraz alebo hodnota daná tými predošlými primitívami, ktoré pokrývali daný pixel),
- hodnota reprezentovaná aktuálne vkladávaným primitívom.

Nová hodnota sa môže určiť rôznym spôsobom, my sme uvažovali tri rôzne metódy skladania hodnôt (a teda skladania viacerých prekrývajúcich sa primitív):

- prepis pôvodnej hodnoty novou hodnotou,
- pripočítanie (akumulácia) novej hodnoty k pôvodnej hodnote,
- agregácia oboch hodnôt ich spriemerovaním.

Tieto tri metódy sú ilustrované na obr. 1 – prepis je vľavo, akumulácia je v strede a spriemerovanie je vpravo.



Obr. 1 Ilustrácia skladania dvoch primitív.

Pri prepisovaní je primitívum priamo vkladané do obrazu – jeho farba priamo prepisuje aktuálnu farbu relevantných pixelov. Výsledok takéhoto skladania je najviac farebný, obrazy pôsobia živším dojmom. Z uvedených troch metód distribúcia pravdepodobnosti generovania rozličných farieb je najrovnomernejšia.

Akumulácia je realizovaná ako prostý súčet hodnôt – výsledná hodnota nie je nižšia ako žiadna z hodnôt na vstupe. Táto metóda obmedzuje počet primitív, ktoré sa môžu prekrývať – pri väčšom počte naakumulovaná hodnota nesená pixelom presiahne maximálne povolenú hodnotu (je to možné riešiť trebárs operáciou modulo). Metóda preferuje jasnejšie farby (reprezentované vyššími hodnotami).

Pri spriemerovaní hodnôt dochádza k vyhladzovaniu a zároveň ostáva viditeľné segmentovanie obrazu pomocou skladania geometrických primitív. Preferované sú farby reprezentované strednými hodnotami, s potlačením extrémnych hodnôt.

Nakoľko akumulácia metóda v úvodných pokusoch buď generovala chybu alebo pri zavedení obmedzení viedla na horšie výsledky ako ostatné dve metódy, tak v následných experimentoch už nebola využívaná.

C. Štruktúra algoritmu

Po prvotných pokusoch sme sa rozhodli použiť reprezentáciu priamo pomocou RGB matice. Generovanie začína vždy z prázdneho obrazu (reprezentovaného čiernou plochou). Do tohto obrazu sú postupne pridávané náhodne generované geometrické primitíva, pričom pridávané primitívum je skladané s aktuálnym obrazom (a teda primitívami pridanými v predošlých

etapách) buď metódou prepisu alebo spriemerovania. Tomu bola prispôsobená aj voľba genetických operátorov – použitý bol iba mutačný operátor, pridávajúci náhodne generované primitívum do aktuálneho obrazu.

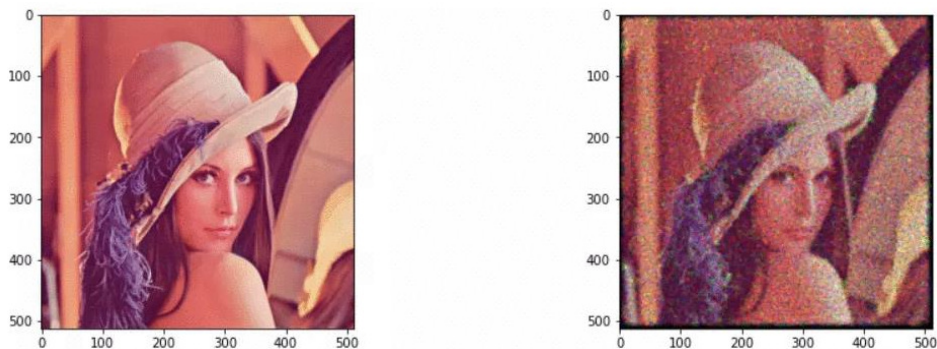
Podobnosť generovaných obrazov voči cieľovému obrazu je použitá ako vhodnosť týchto generovaných obrazov. Je určovaná na úrovni jednotlivých pixelov, pričom je úmerná rozdielu hodnoty pixelu generovaného a cieľového obrazu.

Pri výbere rodiča sa vyberá stále zo všetkých jedincov v populácii a zároveň najlepšieho jedinca predošlej generácie. Tým sa zabezpečí, že ak dôjde u každého jedinca v generácii k zhoršeniu vhodnosti, stále bude prístup k najlepšiemu jedincovi z predošlej generácie. Tento elitistický prístup zabezpečí monotónnosť zmeny vhodnosti najlepšieho jedinca v populácii, pretože počas evolúcie nebude dochádzať k zhoršovaniu najlepšieho jedinca.

III. VYBRANÉ EXPERIMENTY S GENEROVANÍM OBRAZOV

A. Lena - úvodný test

Jedným z faktorov, ktoré ovplyvňujú výsledný generovaný obraz, je veľkosť generovaných mnohoúhelníkových primitív. Tá je síce volená náhodne, avšak z určitého zadaného rozsahu. Vhodnou voľbou tohto rozsahu je možné ovplyvniť ako presnosť generovania tak aj výsledný vizuálny dojem. Pri snahe dosiahnuť stav, aby sa vygenerovaný obraz čo najviac podobal cieľovému obrazu, bolo nutné používať čo najmenšie primitíva (menšie primitíva lepšie vytvárajú detaily). To však viedlo na stav, keď výsledný obraz pripomínal viac šum ako skladanie (mozaiku) geometrických primitív (obr. 2) pri použití štandardného testovacieho obrazu Lena.



Obr. 2 Lena – cieľový (vľavo) a generovaný (vpravo) obraz

Preto v ďalšom experimente sme preferovali používanie väčších geometrických primitív. Výsledná presnosť (pixel voči pixelu medzi cieľovým a generovaným obrazom) síce poklesla, avšak subjektívne bol generovaný obraz zaujímavejší.

B. Mona Lisa

Ako cieľový obraz sme použili reprodukciu obrazu Mona Lisa (aktuálne umiestneného v parížskom Louvri) – známeho obrazu Leonarda Da Vinci. Použitý cieľový obraz (obr. 3) mal rozmery 957x540 pixelov. Tento obraz sme použili kvôli dojmu nízkeho kontrastu, gradientu a tmavým farbám – podmienkam, ktoré v našich predbežných experimentoch viedli k lepším výsledkom.

Pri všetkých pokusoch maximálny rozsah veľkosti geometrických primitív v nejakej osi bol limitovaný tretinou veľkosti cieľového obrazu v danej osi, pretože na obraze jednotlivé rovnomeré plochy nie sú väčšie ako tretina rozlíšenia.

Vzhľadom na výpočtovú zložitosť sme použili zmenšený rozmer obrazov (na jednu tretinu) ako kompromis medzi potrebným časom výpočtu a zachovaním hodnotiteľnej kvality generovaného obrazu. Všetky nasledujúce obrazy boli generované počas 300 000 generácií evolučného procesu.



Obr. 3 Leonardo Da Vinci, Mona Lisa (zdroj: Wikimedia Commons)

Prvý vygenerovaný obraz bol vytváraný trojuholníkovými geometrickými primitívami. Pre skladanie primitív bola použitá metóda priameho prepisovania primitív. Obraz (obr. 4) pôsobí viac farebne, trojuholníky sú viditeľné kvôli použitiu priameho zápisu do obrazu a medzi primitívami je vidieť členitý prechod.



Obr. 4 Generovaná Mona Lisa – trojuholníkové primitíva s prepisom

Druhý obraz bol vytváraný pomocou obdĺžnikových geometrických primitív (obr. 5). Spôsob skladania primitív bol rovnaký ako v predchádzajúcom prípade.



Obr. 5 Generovaná Mona Lisa – obdĺžnikové primitíva s prepisom

Tretí vygenerovaný obraz (obr. 6) bol vytváraný pomocou trojuholníkových primitív, avšak na rozdiel od obr. 4 bola pre skladanie primitív použitá metóda spriemerovania, keď hodnoty pixelov, pokrytých vkladným primitívom, sa nahradia priemerom pôvodnej hodnoty a hodnoty nesenej daným primitívom. Generovaný obraz je menej sfarbený pretože spriemerované hodnoty sú menej výrazné. Navyše nie je viditeľný členitý prechod.



Obr. 6 Generovaná Mona Lisa – trojuholníkové primitíva so spriemerovaním

Štvrtý vygenerovaný obraz (obr. 7) bol vytváraný pridávaním primitív obdĺžnikového tvaru rovnako ako obr. 5, avšak na rozdiel od neho bola pre skladanie primitív použitá metóda spriemerovania. Opäť výsledkom je menej farebný obraz.



Obr. 7 Generovaná Mona Lisa – obdĺžnikové primitíva so spriemerovaním

V experimente vygenerované obrazy dosiahli výslednú vhodnosť (na základe porovnania s cieľovým obrazom) podľa nasledujúcej tabuľky (tab. 1):

Tabuľka 1
Dosiiahnuté vhodnosti generovaných obrazov.

	Skladanie prepisom	Skladanie spriemerovaním
Trojuholníkové primitíva	0.702	0.746
Obdĺžnikové primitíva	0.738	0.761

Z tabuľky je zrejme, že z hľadiska presnosti voči cieľovému obrazu je vhodnejšie použiť geometrické primitíva obdĺžnikového tvaru a pre skladanie jednotlivých prekrývajúcich sa primitív do výsledného obrazu je vhodnejšie preferovať metódu skladania spriemerovaním.

Presnosť reprodukcie cieľového obrazu sa vo všeobecnosti prejavuje v lepších zobrazeniach detailov – avšak tento vzťah nie je absolútnym. Dobrým príkladom je oblasť úst (vykreslenie je lepšie pri obdĺžnikových primitívach) alebo oblasť ľavého oka (vykreslenie je lepšie pri trojuholníkových primitívach).

Navyše príťažlivosť takto generovaných obrazov (vlastne rôznych aproximácií pôvodného cieľového obrazu) je subjektívnou kategóriou. Generované obrazy sme v rámci ankety ukázali respondentom, pričom mali vyjadriť svoju preferenciu jedného voči ostatným. Prieskumu sa zúčastnilo 73 respondentov, pričom väčšina preferovala obraz, ktorý nemal najlepšiu vhodnosť (tab. 2).

Tabuľka 2

Počty respondentov preferujúcich v ankete daný obraz (uvedené sú preferencie iba vyššie zobrazených generovaných obrazov – súčet preferencií je preto menší ako celkový počet účastníkov ankety).

	Skladanie prepisom	Skladanie spriemerovaním
Trojuholníkové primitíva	14	36
Obdĺžnikové primitíva	5	3

IV. ZÁVER

Vykonané experimenty potvrdili, že aproximácia obrazov pomocou skladania prekrývajúcich sa geometrických primitív dokáže vyprodukovať obrazy, ktoré sa ľuďom môžu zdať zaujímavé. Použitie evolučného algoritmu pre tento cieľ sa ukázalo životaschopné. Aj keď oproti pôvodnému obrazu dochádza k strate presnosti, nepresné zobrazenie (resp. často iba naznačenie) detailov je vyvážené zaujímavým segmentovaním obrazu do množiny prekrývajúcich sa plôch. Výsledný dojem z generovaného obrazu je pritom silne subjektívnym – niekto môže preferovať väčšiu a niekto zase menšiu farebnosť, niekomu môžu trojuholníkové segmenty pripadať príťažlivejšie (vytvárajú dojem menšej usporiadanosti než segmenty obdĺžnikové), niekto uprednostní hrubšiu aproximáciu cieľového obrazu pred jemným znázornením detailov. Vhodný výber parametrov generovania a experimentovanie s nimi umožnia generovanie subjektívne príťažlivých verzií cieľových obrazov.

LITERATÚRA

- [1] P. Stubbings, TSP Art (Travelling Salesman Problem Art). Dostupné: <http://parasec.net/blog/tsp-art/> (6.5.2022).
- [2] Image evolution. Dostupné: <https://alteredqualia.com/visualization/evolve/> (6.5.2022)
- [3] B. Traversy, Triangula – an iterative algorithm to generate triangulated and polygonal art form images. Dostupné: <https://github.com/RH12503/Triangula> (6.5.2022).
- [4] Y.N. Dauphin, Clojure: Genetic Mona Lisa problem in 250 beautiful lines. Dostupné: <http://npcontemplation.blogspot.com/2009/01/clojure-genetic-mona-lisa-problem-in.html> (6.5.2022)
- [5] M. Mach, *Evolučné algoritmy: prvky a princípy*. Košice, Elfa, 2009.
- [6] T. Hudák, *Evolúcia obrazu použitím evolučného algoritmu*, Bakalárska práca, Fakulta elektrotechniky a informatiky Technická univerzita v Košiciach, 2021.

Supporting the Education of Algorithms and Data Structures Through Educational Games

¹Kristian LUTAK, ²Slavomír ŠIMONÁK

¹Department of Computers and Informatics, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Slovak Republic

²Department of Computers and Informatics, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Slovak Republic

¹kristian.lutak@student.tuke.sk, ²slavomir.simonak@tuke.sk

Abstract — In this work, we explore the importance of learning games. Based on the results of the exploration, we have created the game Universe Conquest, which can be useful when studying the subject "Algorithms and Data Structures". After describing the game itself, the evaluation of the game follows, which took place in the form of a questionnaire. The positive evaluation results show that learning through games is a very attractive and useful way of gaining new knowledge.

Keywords — web, programming, data structures and algorithms, educational games, interactive learning

I. INTRODUCTION

Games have always been, are, and will most likely be an integral part of our lives. Since childhood, we have been learning and getting to know this world through games. They help us grow and mature. We do not perceive them as something complicated, boring, and pointless. The word "game" is definitely associated with something interesting, unusual or even something that gives us a sense of relaxation.

"Games are thus the most ancient and time-honored vehicle for education. They are the original educational technology, the natural one, having received the seal of approval of natural selection" [1]. Written by Chris Crawford in his book *The Art of Computer Game Design*. This statement proves the importance of games in any aspect of education. Over time, games have been used as a teaching tool for all different situations of life. Puzzles to teach logical thinking, math games to improve basic math skills, and even reading games to increase reading ability are all used successfully to teach the essential skills people will need in life. It, therefore, follows logically that the use of computer games is also an effective way of teaching programming.

With the beginning of active technology development, we began to live in two worlds: real and digital. The concept of games has not disappeared but has instead been actively reformatted and developed in the digital world without losing its two main functions: entertainment and education. Educational games are useful even for those who do not want to waste time and strive to learn something new every day. They can choose from a wide range of games for learning foreign languages, sciences, developing creativity or critical thinking.

The use of games in the educational environment is becoming increasingly popular [2]. This brings benefits not only to students but also to teachers. Game-based learning is a great addition to any subject's teaching material. The teacher always faces a difficult task because he needs not only to convey the material in a form that is accessible and understandable to all students but also to engage them and make sure that they understand the material. This task is further complicated by the fact that each student is a unique individual with his own views, qualities, and needs. Any educational game is aimed at facilitating the acquisition of material and reducing the stress of learning. The benefits of educational games have inspired us to do both contribute to their development and help teachers and students in the educational process.

We have created a game called Universe Conquest [3], which helps to explore the algorithms and data structures learned in the Algorithms and Data Structures course in a playful way. This is an educational game designed for entertainment and learning with a simplified game program. The purpose of the game is to attract the attention of students, to make them more interested in the course of the lessons, to help them learn and apply new knowledge and make them sure that the process of learning can be fun.

II. UNIVERSE CONQUEST

The main goal in the development of our game was to make the game fun and at the same time provide players with new knowledge and motivate them to learn [4]. The game works as follows: the player starts playing a game in which he must solve problems. If the player is unable to complete the task, he can use a hint. After each finished task, the player must solve a mini-problem or puzzle in order to open the next level. The performance of a player can be measured by completing tasks: how many times he won and how many times he lost [5]. The process of the game itself can be seen in Fig. 1.

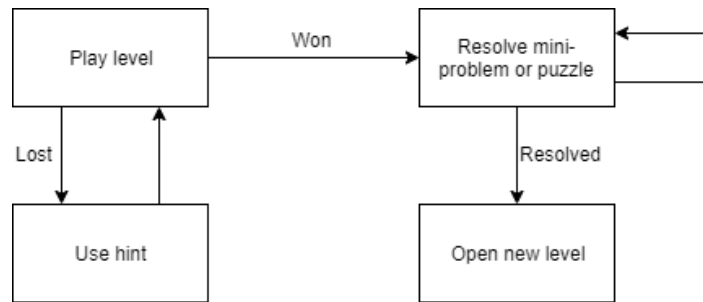


Fig. 1 Game process.

A. The concept of the game

At the center of the game is a player who manipulates various objects during the game. These objects are Lives, Hints, Levels, Secret codes, Time limits, and Data Structure items. At the beginning of the game, the player has a set number of lives that can increase or decrease during the game. The player has the opportunity to choose the level he wants to solve only if he knows a secret code of the chosen planet (level). The player's task is to deal with various problems by manipulating objects that symbolize data structures: the player moves or clicks on these objects. If the player makes a mistake or runs out of time, he loses one life. If he completes all the tasks of the level successfully, he gets one life.

B. World and game levels

According to the typology of games, we classify the game as a quest.

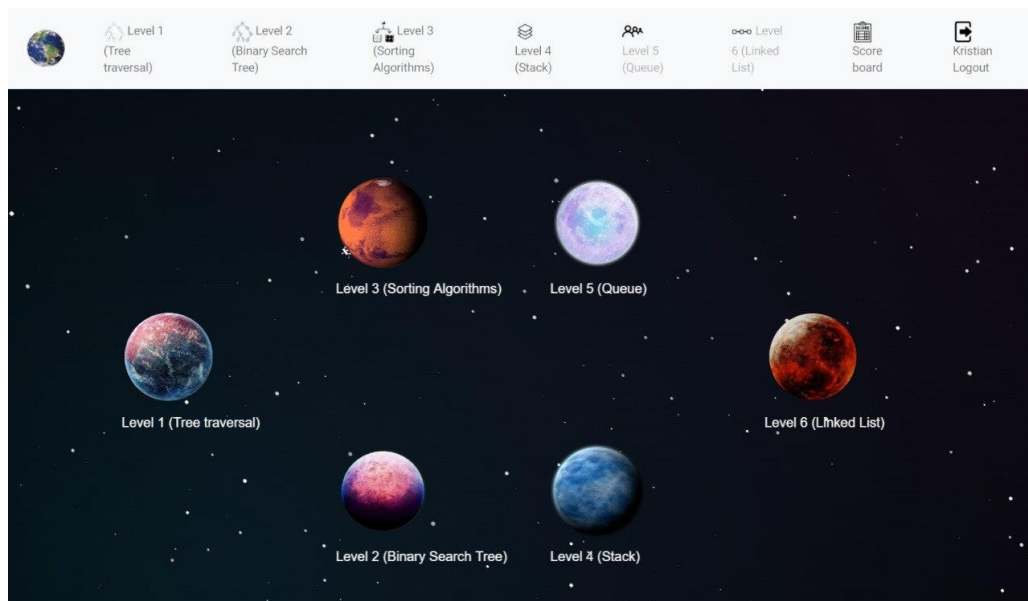


Fig. 2 Game world.

The main character is the player who finds himself in space. In this world (Fig. 2), there are various planets (levels) that the player must successfully overcome. Each level covers different educational areas (binary search trees, sorting algorithms, stack, front, linked list).

Level Tree Traversal (Fig. 3). This level checks whether the player can work with binary trees and especially their transitions: level-order, inorder, preorder, postorder. The player should work with a tree consisting of six nodes. He needs to click on these nodes in the correct order, depending on the type of tree transition. The player has 20 seconds for each task. If the timer expires or the player goes through the tree in the wrong order, he loses one life.

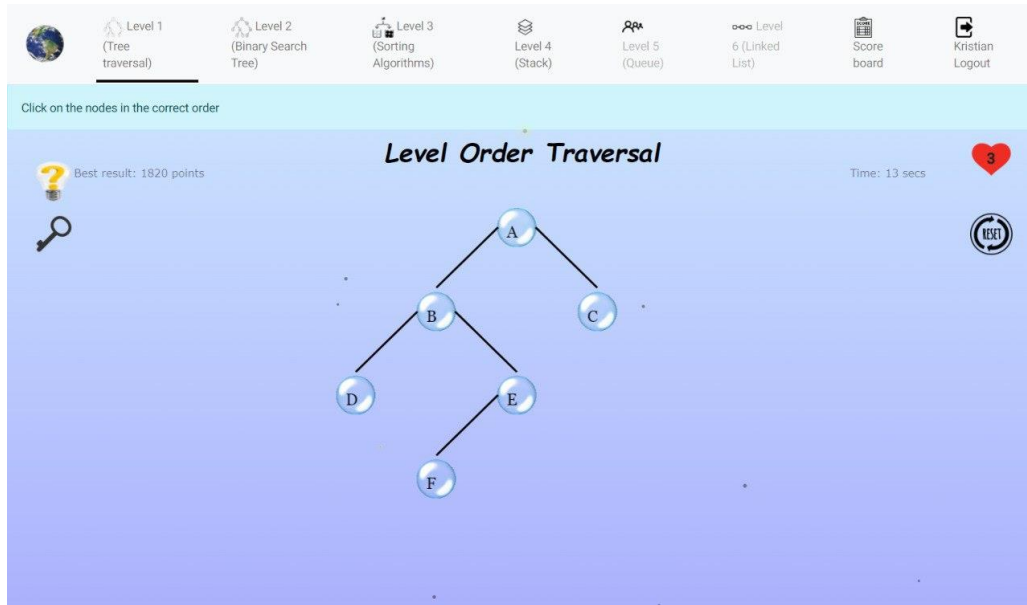


Fig. 3 Tree traversal.

Level Binary Search Tree. This level is similar to the previous Tree Traversal, however here the player should build a binary tree from the available objects. He faces three tasks of different difficulty levels. As in the preceding level, the player has 20 seconds for each task.

Level Sort (Fig. 4). This level contains three tasks for testing knowledge of sorting methods: bubble sorting, selection sorting, and insert sorting. As in the first level of Tree Traversal, the player must click on objects depending on the type of sorting. The player has 100 seconds to complete each sorting method.

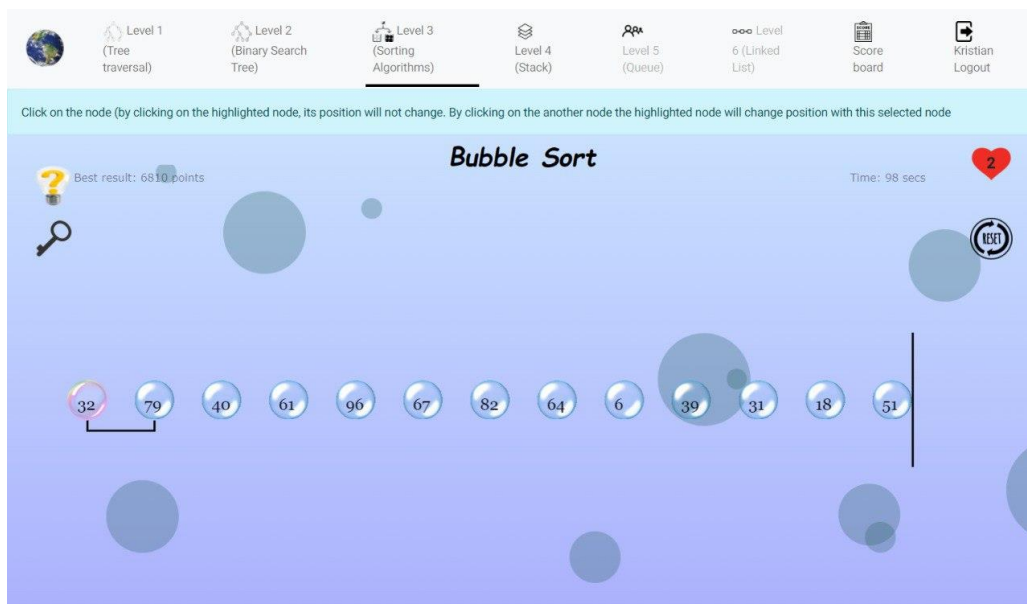


Fig. 4 Sorting algorithms.

Levels Stack and Queue (Fig. 5). These two levels cover two thematic areas (stack and queue). Levels contain three main tasks. Task 1: insert elements into the stack / queue that acquaint players with the Push and Enqueue operations. Task 2: remove elements from the stack / queue that introduce players to the Pop and Dequeue operations. Task 3: sort the elements in the stack / queue according to the level conditions using the Push Pop / Enqueue Dequeue operations. The player has 15 seconds to insert and remove elements and 60 seconds to sort the stack / queue.

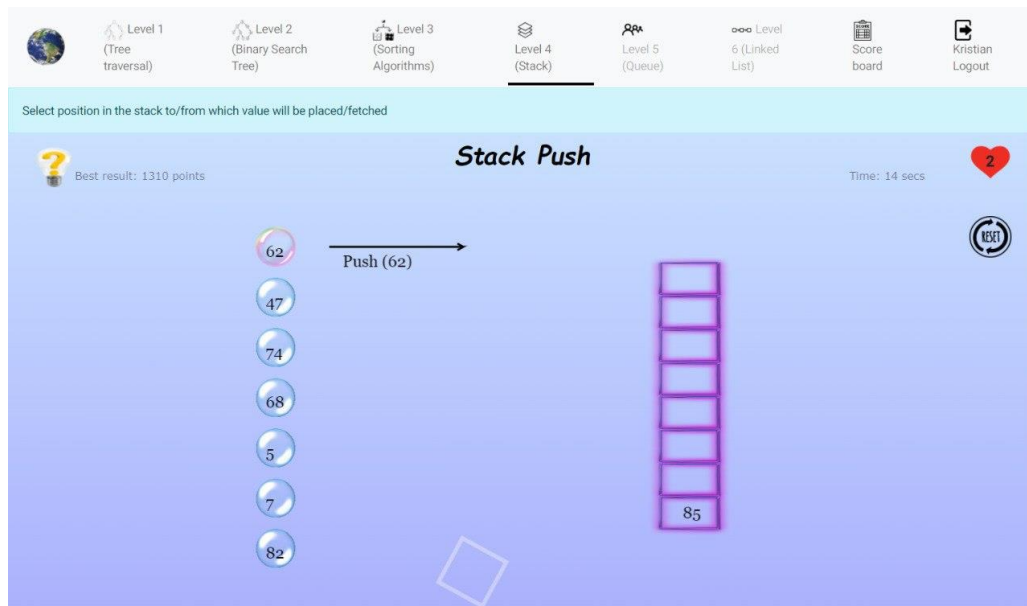


Fig. 5 Stack.

Level Linked List. The LinkedList level contains two tasks. Task 1: insert elements into a linked list. The task imitates the Insert operation. Task 2: remove elements from the linked list. The task resembles the Delete operation. The player has 25 seconds to complete these tasks.

C. Implementation

When creating the game, we decided that it would be a web application written in Angular (frontend) and Node.js Express (backend). We came to a conclusion that it is better not to use any of the game frameworks to draw all the objects and levels, but instead to use pure JavaScript and the HTML Canvas element. The main reason for this decision is the increasing of page loading times, due to amount of predefined codes, which are part of gaming libraries. There are also several additional reasons, which are described in [3].

There are a lot of reasons why we chose Angular [6]. Firstly, it is a cross-platform, powerful, and modern. Secondly, it has a great ecosystem. Thirdly, it provides not only tools but also design patterns to create a sustainable project. When you build an application properly, you will not end up with a mess of classes and methods that are difficult to modify and even harder to test. The code is well-structured, you will quickly understand what is what. Angular already has many tools for creating applications. Directives allow HTML elements to behave dynamically. You can enhance forms with FormControl and provide different validation rules. You can easily send asynchronous HTTP requests of various types. Moreover, you can efficiently set up routing. In Angular, the display and application logic are separate, which greatly cleans and simplifies labeling.

We chose ExpressJS [7] to create the backend API. It allows you to create a robust API quickly and easily. Express is a server framework for the Node.js web application, which is designed specifically for creating single-page, multi-page, and hybrid web applications. Express is a pre-built Node.js framework that allows you to develop faster and smarter server-side web applications. Some of its features are accessibility, minimalism, versatility, and usability. Basically, whatever Bootstrap does for HTML / CSS and responsive web design, Express does for Node.js. Express is now the most popular Node.js framework. The game itself can be found at <https://dp-educational-game.herokuapp.com>.

D. Evaluation

After the game was finished, 20 people were asked to try it out and share their opinions. All 20 applicants were students of Technical University of Košice who had already completed the Algorithms and Data Structures course. They provided the feedback about the effectiveness of the game leaning on their experience.

Testing took place in the form of a questionnaire that respondents filled out after playing the game. The questionnaire consists of 14 questions, the purpose of which is to find out the opinion of the players about the game itself, its advantages, and disadvantages. We also asked if this game is suitable for the educational process, what the motivation of the players was and what difficulties they encountered when completing the levels.

The game attracts the attention of 75% of respondents. Also, learning through game is considered very attractive. More than half of the respondents, that is, 15 people, admitted that this game would inspire them to take an active participation during the lessons and boost their knowledge. The majority of the respondents find the game fun and rewarding in terms of learning and practicing algorithms. We asked the respondents which topic visualization was the most useful to them. According to the votes it was sorting algorithms. On the other hand, 70% of respondents also reported, they used a hint. This may indicate that some tasks were too complicated for them or were not expressed clearly enough.

In general, the game received a positive assessment, which means it successfully fulfilled its role. It introduced the players to working with the main points of the subject Algorithms and Data Structures which are binary trees, sorting algorithms, stack, queue, and linked list.

III. RELATED WORK

Application of educational games in the field of teaching algorithms and data structures is extensively studied presently. In [8] authors proposed a model integrating algorithm visualization and gamification into learning instructions. They have found, that it can significantly improve the knowledge in the field of sorting algorithms. A collaborative game designed to help students in learning constraint satisfaction algorithms is described in [9]. An evaluation study confirmed promising results in the field of motivation, engagement and interest. An educational game intended to improve understanding of the stack data structure is presented in [10]. The game addresses several teaching aspects, like concepts, application and implementation. And the evaluation of the game disclosed statistically significant learning profits. An interesting study on evaluation of games for computing education can be found in [11].

The game Universe Conquest can be compared to a similar game called IT Adventure implemented by Nagyová [12]. These two games have their similarities and differences. Both of them have the same aim - to help students master and check their understanding of the subject Algorithms and Data Structures in a playful way. The main similarity is the structure of the game. The levels are mainly focused on the primary topics of the subject which are sorting algorithms, binary trees, stack, queue, and linked lists. The player has 3 lives at the beginning of the game. The number of lives can increase or decrease. If the user faces some difficulties during passing the level he can use a hint. Both games have an interesting and attractive design, even though the one sends a player to a laboratory, while another one makes him travel to space.

The major difference is in the games' implementation. The IT Adventure is implemented as a desktop application. To play it, the user has to download a lot of files and have one of these operating systems - Linux or Windows. The game Universe Conquest is implemented as a web application. To play the game, it is enough to have any modern browser installed on any type of device (smartphones, tablets, notebooks, etc.). Both games challenge the player and make passing the levels more intense and difficult. In IT Adventure you can meet various enemies that prevent you from passing the level, while Universe Conquest has time limits which make a player afraid of running out of time and failing. Moreover, in the Universe Conquest, there is a space for competition between users, because it contains a scoreboard with the results of the top players.

Both games were tested by students of Technical University of Košice that passed through the subject Algorithms and Data Structures during their studies. After playing the games they completed questionnaires. According to them, 84% of respondents [13] admitted that IT Adventure got their attention, while 75% were satisfied with Universe Conquest. Generally, games met their requirements and expectations, however, the results of the questionnaires show us that there is still a space for improvement.

IV. CONCLUSION

The use of computer learning games in education is a current trend in connection with the active introduction of information and communication technologies in the educational process. Educational games have strong pedagogical potential. Through activities, they enable learners to acquire the necessary knowledge, train practical skills, and control the assimilation of the material. Games also make it easier to visualize the principles of individual algorithms. They are a means of activation of cognitive activity and contribution to the intellectual development of students.

The game called Universe Conquest was created as a web application and designed to be fun, easy to use, and provide players with new knowledge. The game introduces players to the algorithms they have mastered in the course Algorithms and Data Structures in an entertaining way. Of course, the application does not replace the programming itself within this subject but helps players understand how certain principles of algorithms and data structures work. The game can also serve as an aid in learning this subject.

The game received a very positive evaluation, so it is definitely worth of thinking about the introduction of various types of educational games into the educational process. The respondents in the questionnaire agreed that thanks to this game the educational process would be more attractive and interesting. Of course, creating a game is a very long process that can take months, or years to be introduced to the public. Therefore, there is still some room for development and improvement in our game. At least based on the feedback from the players, it would be good to add new levels to the game and some enemies that will interfere with the player. We can conclude that Universe Conquest can become a useful part of education.

REFERENCES

- [1] C. CRAWFORD, *The Art of Computer Game Design*. Washington State University, 1997.
- [2] S. PAPADAKIS, *The use of computer games in classroom environment*. International Journal of Teaching and Case Studies (IJTCS), Vol. 9, No. 1, 2018.
- [3] K. LUTAK, *Application of Educational Games in Teaching Algorithms and Data Structures*. Master's thesis, Technical University of Košice, 2022.
- [4] S. S. SHABANAH, J. X. CHEN, H. WECHSLER, D. CARR, E. WEGMAN, *Designing Computer Games to Teach Algorithms*. Seventh International Conference on Information Technology: New Generations, 2010, pp. 1119-1126.
- [5] T. BAIBAK, R. AGRAWAL, *Programming Games to Learn Algorithms*. 2007 Annual Conference & Exposition, Honolulu, Hawaii, 2007.
- [6] A. K. OHLUKYAN, *Overview of the Angular Framework: Pros and Cons*. OMEGA SCIENCE, 2020.
- [7] A. MARDAN, *Express.js Guide: The Comprehensive Book on Express.js*. 2014.
- [8] A. YOHANNIS, Y. PRABOWO, *Sort Attack: Visualization and Gamification of Sorting Algorithm Learning*, 7th International Conference on Games and Virtual Worlds for Serious Applications (VS-Games), 2015, pp. 1-8.
- [9] F. GRIVOKOSTOPOULOU, I. PERIKOS, I. HATZILYGEROUDIS, *A Collaborative Game for Learning Algorithms*, Proceedings of the 9th International Conference on Computer Supported Education (CSEDU 2017) - Volume 1, pp. 543-549.
- [10] D. DICHEVA, A. HODGE, *Active Learning through Game Play in a Data Structures Course*, Proceedings of SIGCSE 2018, Baltimore, MD, USA, 2018, pp. 834-839.
- [11] G. PETRI, C. G. VON WANGENHEIM, *How games for computing education are evaluated? A systematic literature review*, Computers & Education, Vol. 107, 2017, pp. 68-90.
- [12] K. NAGYOVÁ, S. ŠIMOŇÁK, *Application of Educational Computer Games in Algorithms and Data Structures Courses*. Electrical Engineering and Informatics XI, Proceedings of the Faculty of Electrical Engineering and Informatics of the Technical University of Košice, 2020.
- [13] K. NAGYOVÁ, *Possibilities of Application of Educational Games in Algorithms and Data Structures Courses*. Master's thesis, Technical University of Košice, 2020.

Technologická podpora logopedickej terapie v slovenčine

Stanislav ONDÁŠ, Lea ŠIMČÍKOVÁ

Katedra elektroniky a multimediálnych telekomunikácií, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovensko

¹stanislav.ondas@tuke.sk, ²lea.simcikova@tuke.sk

Abstrakt — Témou tejto predloženej práce je problematika použitia smart technológií v oblasti detskej logopedickej terapie a rehabilitácie. V práci sa zameriame na opis vývoja reči, metodiku logopedickej diagnostiky a terapie u detí a rôzne poruchy reči. Na základe analýzy logopedických metódik a dostupných nástrojov pre podporu logopedickej terapie sme navrhli pilotnú verziu podporného nástroja pre detskú logopedickú terapiu. Táto webová aplikácia je zameraná na nácvik správnej výslovnosti jednotlivých hlások v slovenskom jazyku s využitím DTW algoritmu na určenie miery podobnosti vysloveného slova so vzorovým slovom so správnou výslovnosťou. Webová aplikácia je realizovaná hravou formou, obsahuje obrázky a nahrávky so správnou výslovnosťou slov, ktoré obsahujú problematickú hlásku. V rámci vývoja aplikácie vznikla aj databáza vzorov slov (celkovo 189 slov) pre nácvik jednotlivých problémových hlások v slovenčine.

Kľúčové slová — logopédia, webová aplikácia, logopedická terapia, nácvik výslovnosti

I. ÚVOD

Dnes, v dobe veľkého technologického pokroku sa inovuje vo všetkých oblastiach nášho života a to aj vrátane zdravotníctva. Výrazný pokrok v tomto odvetví sme mohli postrehnúť najmä kvôli celosvetovej pandémie COVID-19, kedy sa mnoho úkonov realizovalo vďaka internetu a moderným technológiám na diaľku. Pomocou internetu sa ľudia dokázali z pohodlia a bezpečia domovov objednať na vyšetrenie alebo požiadať o niektoré zdravotné záznamy, čo prispievalo k zvládnutiu pandémie COVID-19 avšak zároveň zefektívnilo fungovanie mnohých oblastí aj v post-covidovom období. Je možné konštatovať, že aj vďaka pandémie koronavírusu sa zrýchlil vývoj a nástup mnohých online nástrojov pre podporu telemedicíny, online diagnostiky a terapie v mnohých medicínskych oblastiach. Aj keď sa často jedná iba o substitučné prostriedky, ktoré nie vždy ideálne nahrádzajú fyzickú diagnostiku a terapiu, takto navrhnuté prostriedky môžu zlepšiť dostupnosť zdravotnej starostlivosti a tiež priaznivo ovplyvniť skorú diagnostiku a intervenciu.

Pandémia však negatívne zasiahla aj úroveň komunikačných schopností detí naprieč celým svetom. Reč a komunikácia je pre naše životy veľmi dôležitá a vyvíja sa už v skorom veku, a v súčasnosti môže mnoho detí trpieť poruchou reči v dôsledku nedostatočnej komunikácie a socializácie. Je ale zrejmé, že poruchy reči sa bežne vyskytujú v ľudskej populácii a ich dôsledkami sú napr. aj poruchy komunikácie, problémy so vzdelávaním a zaradením sa do spoločnosti. Zároveň je možné konštatovať, že v oblasti logopédie je stále nedostatok podporných nástrojov, ktoré by umožňovali nácvik správnej výslovnosti v domácom, komfortnom prostredí a atraktívnym spôsobom pomocou moderných technológií, čo špeciálne platí pre reálne slovenského jazyka.

Na základe predchádzajúcich skúsenosti z vývoja aplikácií pre detskú rečovú audiometriu (viď [9], [10]) a kontaktu s komunitou nepočujúcich ako aj s ambulanciou klinickej logopédie Mgr. Oriškovej v Košiciach sme sa začali zaoberať vývojom webovej aplikácie pre podporu detskej logopedickej terapie. Cieľom je priniesť jednoduchý nástroj pre realizáciu nácviku výslovnosti v domácom prostredí, ktorý by bol atraktívny pre detského používateľa.

II. LOGOPÉDIA A LOGOPEDICKÁ TERAPIA

A. Logopédia

Logopédia patrí medzi intervenčné metódy alebo služby, ktoré napomáhajú vyhodnocovať a liečiť komunikačné problémy a poruchy reči. Taktiež sa zameriava na zlepšovanie rečových schopností a schopnosti porozumieť reči. Logopedickú intervenciu zvyčajne realizuje logopéd alebo audiológ. Terapeutické postupy sú zamerané na nápravu porúch v komunikačnom prejave, a tiež na prevenciu zameranú na rozvoj komunikačných schopností u detí. Logopédia sa začala formovať už v prvej polovici 20. storočia, avšak ešte stále sa vyvíja a formuje. Najčastejšími typmi porúch reči, ktoré si vyžadujú logopedickú terapiu sú afázia, dysfázia, detská apraxia reči, dyzartria, dyslália a koktavosť [1]-[4].

Afázia je porucha myslenia, ktorá sa prejavuje poruchou schopnosti vyjadrovať sa rečou a tiež poruchou porozumenia vnímaných slov. Dysfázia je narušením komunikačných schopností, ktoré sa prejavuje neschopnosťou verbálne komunikovať. Apraxia reči je porucha, ktorá spôsobuje ťažkosti v plánovaní reči, kde je narušená komunikácia medzi mozgom a artikulačnými svalmi. Veľmi podobnou poruchou je dyzartria, ktorej hlavnými prejavmi je zhoršená zrozumiteľnosť reči. Dyslália, ako najčastejšia porucha reči, spôsobuje chyby artikulácie, ktoré spôsobujú nesprávnu výslovnosť hlások, resp. ich vynechávanie alebo zamieňanie.

B. Technologická podpora v logopédii

V dnešnej dobe je veľmi málo odvetví, ktoré sú novou technológiou nedotknuté. S primeranou mierou využitia a nasmerovaním správnym smerom týchto technológií, môžu byť prospešné a to nie len v rečovej terapii. Podľa amerického ministerstva zdravotníctva a sociálnych služieb dnes trpí poruchou reči až 8% detí, vo veku od 3 do 17 rokov, pričom tieto poruchy súvisia s prehltním, hlasom, jazykom alebo samotnou rečou. Mnohé z týchto detí absolvujú v škole logopedickú terapiu, niektoré deti sa zúčastňujú online terapie alebo chodia na osobné stretnutia. Komunikačné nedostatky dokážu ovplyvniť životy detí v rôznych aspektoch. Môže to byť ich výkon v triede, schopnosť socializovať sa či už v školských zariadeniach alebo mimo nich, taktiež aj s rodinnými príslušníkmi. Pri nedostatočnom vyjadrovaní pocitov, túžob a potrieb je možné, že to bude viesť k behaviorálnym problémom a môže to postihnúť duševné zdravie a kvalitu života jednotlivca [5].

Najčastejším spôsobom zapojenia moderných komunikačných technológií do terapeutického logopedického procesu je tzv. teleterapia resp. „telepractice“. V tomto prípade sa jedná o tzv. terapiu na diaľku, kde terapeut a pacient spolu interagujú cez telekomunikačné prostriedky (napr. videohovor). Americká asociácia pre reč, jazyk a sluch definuje teleterapiu alebo telepractice, ako aplikáciu telekomunikačnej technológie, ktorá poskytuje profesionálne služby na prepojenie na diaľku medzi lekárom a klientom alebo medzi lekárom a lekárom, ktorej účelom je hodnotenie, intervencia alebo konzultácia. Tento model terapie umožňuje široké spektrum online využitia medzi terapeutom a pacientom.

Terapeutické sedenia môžu byť so súhlasom oboch strán nahrávané, upravované a ukladané. Neskôr si tak terapeut, pacient alebo rodič či učiteľ dokážu pozrieť záznam, ak k nemu majú prístup. Na základe opakovaného záznamu si môže dieťa s poruchou reči precvičovať to, čo sa naučil s terapeutom a zdokonaľovať sa. Prepracované aplikácie a webové programy sú považované za užitočné na zapojenie pacientov a dosiahnutie tak terapeutických výsledkov inými a novými spôsobmi.

Teleterapia okrem iného uľahčuje logopédom prácu, pretože ak pracujú napríklad pre viacero škôl, tak teleterapia im šetrí čas tým, že nemusia cestovať. V dôsledku celosvetovej pandémie COVID-19 sa online logopédia v čase sociálneho dištancovania a diaľkového vzdelávania veľmi zrýchlila, avšak v zahraničí sa trend k teleterapii vyvíja už mnoho rokov. Pri takomto druhu terapie je poskytovaná rovnaká starostlivosť ako pri osobnej terapii. Jediný rozdiel je v tom, že deti komunikujú s logopédom cez video obrazovku. Teleterapiu využíva pre svoje deti stále viac a viac rodičov a taktiež ju poskytuje viacero škôl [6],[7]. Teleterapia v logopédii na Slovensku je poskytovaná napr. Adeli medicínskym centrom (<https://sk.adelicenter.eu/logopedia-online/>).

Hlavnou nevýhodou teleterapie je nevyhnutná prítomnosť terapeuta (logopéda, audiológa) a tým aj obmedzené množstvo pacientov, ktorým môže byť takáto terapia poskytnutá. V logopédii, je ale veľmi dôležité opakované precvičovanie výslovnosti, ktoré sa už môže realizovať bez prítomnosti odborného personálu, pričom ten môže byť nahradený rodičom alebo

učiteľom. V prípade starších detí, by toto precvičovanie mohlo byť realizované aj samostatne, teda bez ďalšej osoby, v prípade využitia vhodných technologických prostriedkov. Tieto podporné nástroje pre oblasť logopédie môžu zohrať dôležitú rolu v procese nácviku výslovnosti v domácom prostredí a tak doplniť klasickú, prezenčnú, logopedickú terapiu, resp. teleterapiu.

Oblasť logopédie je jazykovo-závislá. Pre každý cieľový jazyk je nevyhnutné vytvoriť sadu slov a metodiku pre nácvik výslovnosti jednotlivých hlások v danom jazyku. Pre slovenčinu je možné nájsť niekoľko logopedických aplikácií, ktoré sú voľne dostupné pre rôzne platformy (Android, iOS, MS Windows). Spoločným menovateľom týchto, veľmi užitočných aplikácií je, že nevyužívajú algoritmy evaluácie výslovnosti na báze rečových technológií.

Naším cieľom bolo predstaviť pilotnú verziu aplikácie pre podporu logopedickej terapie v podobe nácviku správnej výslovnosti s využitím algoritmov rozpoznávania reči pre výpočet podobnosti vyslovených slov s ich správne vyslovenými vzormi. Automatické porovnanie vzoru s vysloveným slovom umožňuje vyhodnotiť úroveň vyslovenia problematických hlások a rozhodnúť či je potrebné dané slovo ďalej precvičovať, alebo je možné sa presunúť na ďalšie.

Nesmieme zabúdať, že veľmi dôležitým aspektom pri technológiách v logopédii je rovnováha. Je nutné, aby dieťa strávilo napríklad pri obrazovke počítača alebo tabletu len takú dobu, ako to je nevyhnutné. Pretože ako mnohé výskumy ukazujú, mladšie dieťa môže mať potencionálne oneskorenie reči, práve z dôvodu častého používania týchto zariadení [8].

III. WEBOVÁ APLIKÁCIA PRE NÁCVIK SPRÁVNEJ VÝSLOVNOSTI

Webová aplikácia pre nácvik správnej výslovnosti je určená na nácvik výslovnosti jednotlivých hlások v slovenčine, pričom tieto sú súčasťou sady slov, v ktorých sa hlásky vyskytujú na začiatku, v strede a na konci slova. Webová aplikácia obsahuje celkovo 189 slov, ktoré sú roztriedené do skupín, podľa toho o akú hlásku sa jedná a zároveň podľa toho, kde sa v slove sa problematická hláska nachádza. Jedná sa o 21 hlások, pričom na precvičovanie každej hlásky je pripravených 9 slov.

Na tvorbu webovej aplikácie sme použili štandardné jazyky (HTML, CSS, JavaScript a PHP). Evaluačné skripty pre porovnanie vzorového slova s vysloveným na báze algoritmu DTW (Dynamic Time Warping) boli realizované v Matlabe. Pomocou HTML sme definovali štruktúru webovej stránky, CSS a Bootstrap zabezpečujú jednoduchý a prehľadný vzhľad. JavaScript bol použitý na zabezpečenie funkčnosti, napríklad nahrávania/prehrávania audio súboru a podobne. Pomocou PHP webová stránka komunikuje so serverom a spúšťa porovnávacie skripty.

Na porovnanie zvukových nahrávok, využívame v tejto práci dynamickú deformáciu časovej osi (DTW). DTW je algoritmus, ktorý vypočítava optimálnu dráhu deformácie medzi dvoma zvukovými údajmi. Výstupom tohto algoritmu je vzdialenosť medzi vzorovým slovom a vysloveným slovom. Výpočet vzdialenosti dvoch slov, aj napriek tomu, že ide obsahovo o tie isté slová, je, okrem iného, komplikovaný časovou variabilitou pri produkcii reči. Túto časovú variabilitu je nutné kompenzovať správnym spôsobom. V prípade algoritmu DTW sa jedná o výber vektorov do výpočtu vzdialenosti na základe ich vzájomnej vzdialenosti. Algoritmus hľadá cestu v prehľadávacom priestore (v matici vzdialenosti), tak, aby minimalizoval výslednú vzdialenosť. Alternatívnym prístupom môže byť modelovanie reči pomocou skrytých Markovových modelov, u ktorých sa vykonáva porovnanie pravdepodobností, že pozorovaná sekvencia vektorov mohla byť vygenerovaná daným vektorom.

Vzhľadom na to, že v našej práci nás zaujíma priamo podobnosť dvoch slov, algoritmus DTW nám priamo poskytuje hodnotu vzdialenosti medzi nimi. Pre parametrizáciu vstupného slova ako aj vzoru slov v pamäti sú použité lineárne predikčné kepstrálne koeficienty (LPCC).

Obrázky Obr. 1 – 4. predstavujú základné prvky hlavného rozhrania aplikácie.

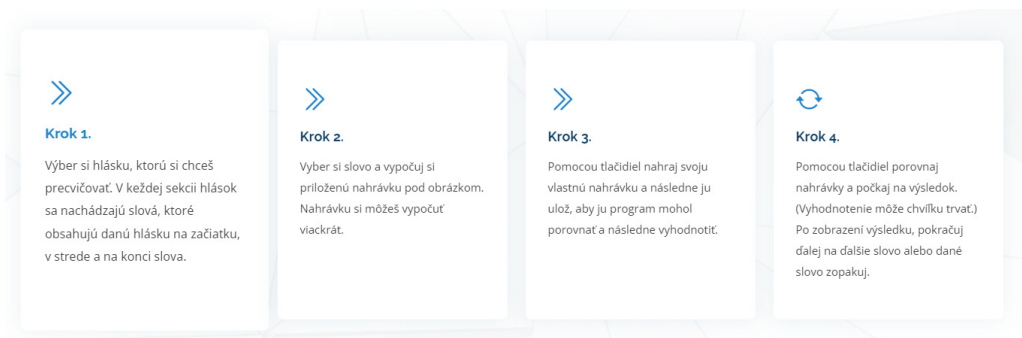


Obr. 1 Navigačný panel domovskej stránky

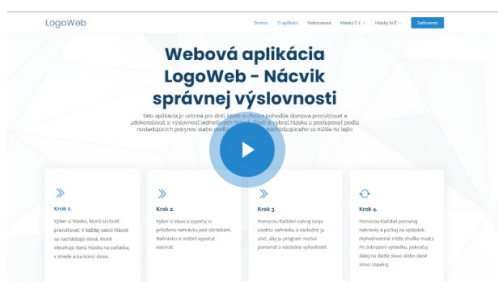
Webová aplikácia LogoWeb – Nácvik správnej výslovnosti

Táto aplikácia je určená pre deti, ktoré si chcú z pohodlia domova precvičovať a zdokonaľovať si výslovnosť jednotlivých hlások. Stačí si vybrať hlásku a postupovať podľa nasledujúcich pokynov alebo podľa videonávodu nachádzajúceho sa nižšie na tejto stránke.

Obr. 2 Nadpis a stručná definícia webovej aplikácie



Obr. 3 Ukážka krokov ako postupovať na stránke



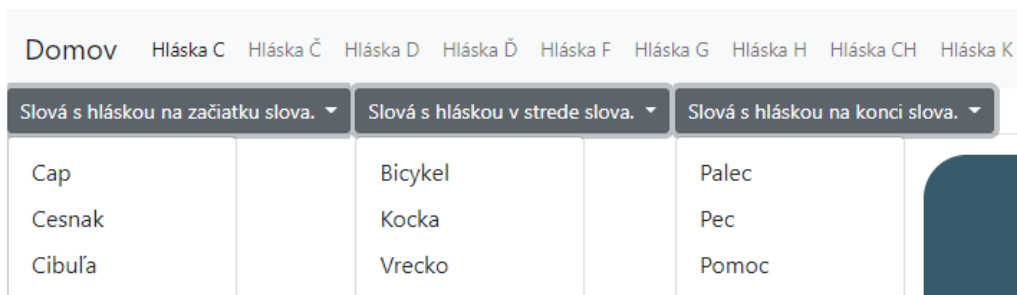
Na tomto videu si môžeš pozrieť názornú ukážku, ako postupovať na tejto webovej aplikácii.

Postup:

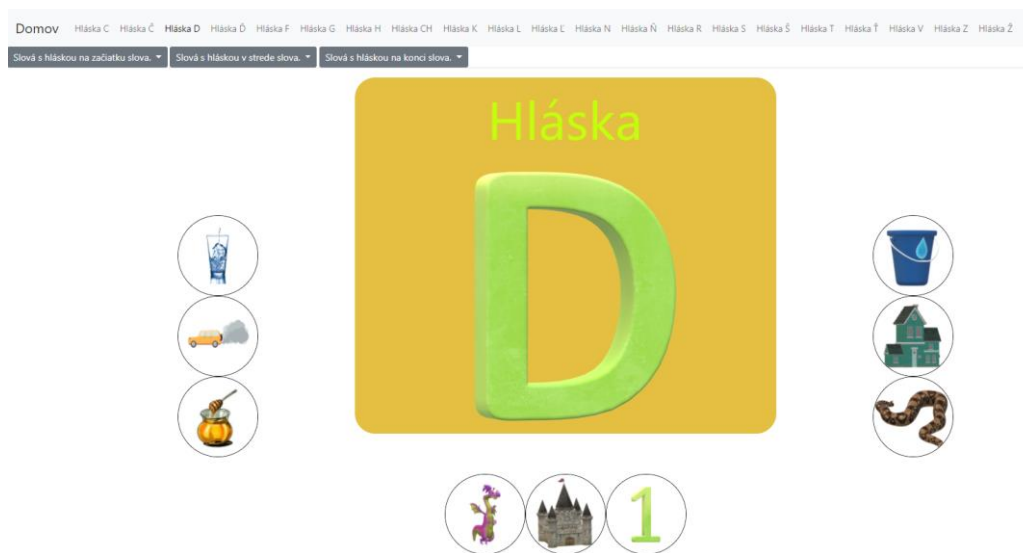
- ✓ Vyber si hlásku, ktorú si chceš alebo potrebuješ precvičovať.
- ✓ Prejdi na dané slovo a vypočuj si priloženú nahrávku.
- ✓ Pomocou tlačidla "Nahrávanie" nahraj svoju nahrávku, ukonči nahrávanie pomocou tlačidla "Koniec nahrávania" a ulož svoju nahrávku pomocou tlačidla "Nahraj", ktoré sa objaví po ukončení nahrávania.
- ✓ Klikni na tlačidlo "Porovnaj", počkaj si na výsledok a pokračuj ďalej.

Obr. 4 Sekcia s videom na domovskej stránke

Ďalšou dôležitou časťou aplikácie je ponuka jednotlivých hlások na precvičovanie a k nim prislúchajúca ponuka slov. Po kliknutí na konkrétnu hlásku resp. slovo je možné začať s jeho precvičovaním. Obr. 6 zobrazuje okno precvičovania hlásky D. V okolí zobrazenia precvičovanej hlásky sa nachádzajú jednotlivé slová.



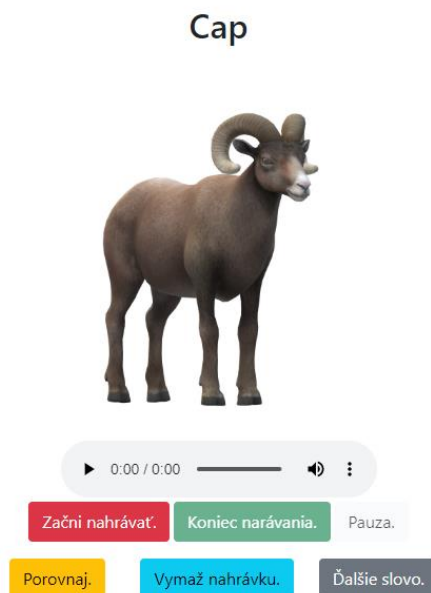
Obr. 5 Rolovateľná ponuka slov hlásky C



Obr. 6 Ukážka webovej stránky hlásky D

Po výbere konkrétneho slova sa zobrazí okno nácviku výslovnosti (Obr. 7.). Stránka jednotlivých slov obsahuje názov a obrázok daného slova, audio prehrávač, panel na nahrávanie a tlačidlá na nahratie nahrávky na server, porovnanie, vymazanie nahrávky a prepojenie na ďalšie slovo.

Po zaznamenaní vysloveného slova je potrebné v pilotnej verzii stlačiť tlačidlo *Porovnaj*, ktoré spustí porovnávanie nahraného a zdrojového zvuku. Oba zvuky sú rozdelené na rámce o dĺžke 25ms a parametrizované pomocou LPCC koeficientov (lineárne predikčné kepstrálne koeficienty). Následne je pomocou DTW algoritmu vypočítaná vzdialenosť medzi maticovými reprezentáciami vstupných signálov. Na základe experimentálne nastaveného prahu sa následne vyhodnotí miera podobnosti ako uspokojivá alebo nedostatočná.



Obr. 7 Ukážka webovej stránky slova Cap

Po porovnaní sa zobrazí výsledok, ktorý odkáže užívateľa na ďalšie slovo alebo ho vyzve, aby dané slovo zopakoval. Po nahraní nahrávky je užívateľ schopný nahrávku vymazať pomocou tlačidla 'Vymaž nahrávku'.



Obr. 8 Výpis obrazovky po úspešnom (vľavo) a neúspešnom (vpravo) pokuse nahrávania a porovnávania

IV. ZÁVER

Cieľom predloženej práce bolo naštudovať problematiku používania technológií v oblasti logopedickej terapie a rehabilitácie a vytvoriť pilotnú verziu logopedickej aplikácie, ktorá by umožňovala automatické porovnanie referenčného a vstupného zvuku pre vyhodnotenie správnosti vyslovenia navčičovaného slova. Motiváciou bolo vytvoriť aplikáciu pre nácvik správnej výslovnosti, ktorá by bola atraktívna pre detského používateľa a zároveň by umožňovala nácvik výslovnosti v domácom prostredí bez potreby účasti logopéda alebo pracovníka so špeciálnym vzdelaním v tejto oblasti. Veríme, že nasadenie rečových technológií aj v tejto oblasti môže priniesť benefity vo väčšej motivácii detí podstúpiť logopedickú terapiu, v dostupnosti nástrojov pre nácvik výslovnosti a tým prispieť k dostatočnému rozvinutiu komunikačných schopností detí.

Vzhľadom na to, že táto aplikácia je zatiaľ iba pilotnou verziou v budúcnosti sa budeme sústrediť na jej celkové vylepšenie, vrátane vylepšenia porovnávania slov, zlepšenia ergonomie rozhrania a na implementovanie ďalších logopedických postupov. V rámci ďalšieho vývoja sa zameriame na riadne otestovanie navrhutej aplikácie.

POĎAKOVANIE

Táto práca bola podporená Agentúrou na podporu výskumu a vývoja vďaka projektu APVV SK-TW-21-0002, Vedeckou grantovou agentúrou realizáciou výskumného projektu VEGA 1/0753/20 „Robustné rečové technológie metódami hlbokého učenia“ a VEGA 2/0165/21 „Technológie automatického spracovania reči na pomoc v krízových situáciách“ financovaných z prostriedkov Ministerstva školstva, vedy, výskumu a športu SR.

LITERATÚRA

- [1] Barbara Smith, June 24, 2014, What Is Speech Therapy?. [Online]. [cit. 2021-06-03] Dostupné na internete: <<https://www.parents.com/kids/development/learning-disabilities/what-is-speech-therapy/>>.
- [2] Adrienne Santos-Longhurst, May 9, 2019, What Is Speech Therapy?. [Online]. [cit. 2021-08-03] Dostupné na internete: <<https://www.healthline.com/health/speech-therapy#what-to-expect>>.
- [3] Jiřina Klenková. Grada, 2006. Logopedie. ISBN-8024711109.
- [4] Kejkličková, I. (2016). Vady reči u dětí. Praha: Grada.
- [5] E Leanne Sherred, 2020. How Technology is Redefining Speech-Language Pathology [Online]. [cit. 2021-10-16] Dostupné na internete: <<https://www.emergingedtech.com/2020/12/how-technology-is-redefining-speech-language-pathology/>>.
- [6] Towey M. P. (2012). Speech telepractice: installing a speech therapy upgrade for the 21st century. International journal of telerehabilitation, 4(2), 73–78. <https://doi.org/10.5195/IJT.2012.6112>.
- [7] Leanne Sherred, 2020. How Technology is Redefining Speech-Language Pathology [Online]. [Cit. 2021-10-24] Dostupné na internete: <<https://www.emergingedtech.com/2020/12/how-technology-is-redefining-speech-language-pathology/>>.
- [8] CBS Therapy, 2021. 4 Ways to Implement Technology into Speech Therapy. [Online]. [cit. 2021-10-16] Dostupné na internete: <<https://cbstherapy.com/4-ways-to-implement-technology-into-speech-therapy/>>
- [9] S. Ondas, E. Kiktova, M. Pleva, M. Oravcova, L. Hudak, J. Juhar, and J. Zimmermann, Pediatric Speech Audiometry Web Application for Hearing Detection in the Home Environment. Electronics. 2020; 9(6):994.
- [10] E. Kiktova, J. Zimmermann, S. Ondas, M. Pleva, J. Juhar, and V. Solt esova, The Role of Hearing Screening Using an Audiometry Application in the Education of Children with Hearing Impairment, ICETA 2020 - 18th IEEE International Conference on Emerging eLearning Technologies and Applications, Proceedings, art. no. 9379250, pp. 311- 317.

Algoritmy strojového učenia v identifikácii trolizmu

¹Kristína MACHOVÁ, ²Michal POREZANÝ

^{1,2} Katedra kybernetiky a umelej inteligencie, Fakulta elektrotechniky a informatiky, Technická univerzita Košice

¹kristina.machova@tuke.sk, ²michal.porezany@student.tuke.sk

Abstrakt — Článok sa zaoberá identifikáciou trolov v online komunitách pomocou metód umelej inteligencie. Internet sa stáva čoraz dostupnejším a rozšírenejším. S narastajúcou používateľskou základňou sa častejšie stretávame aj s anti-sociálnym správaním sa používateľov. Jednou z týchto foriem správania je aj trolizmus. Problém trolov sa dostáva do čoraz väčšej pozornosti a s tým spojené návrhy riešení problému odhaľovania trolizmu a jeho regulácie v online priestore. Jedným z riešení je aj identifikácia trolizmu pomocou strojového učenia. Práca pojednáva o správaní trolov na internete a možnostiach, ako a pomocou akých dát a prístupov ich možno odhaliť.

Kľúčové slová — sociálne médiá, trolizmus, identifikácia trolov, strojové učenie, SVM, Naivný Bayes klasifikátor, logistická regresia

I. ÚVOD

Žijeme v informačnej dobe. Informácie na nás siahajú z každej strany. Z printových médií, z televízie či rozhlasu, šírením po vlnách, alebo internete. V posledných rokoch je práve internet najväčším, najrozsiahlejším a najrýchlejším zdrojom informácií, so stále stúpajúcim trendom. S tým je spojený aj prístup k informáciám rôzneho charakteru či už so zdrojom relevantným, alebo nedôveryhodným. Takmer polovica svetovej populácie a väčšina populácie rozvinutých krajín má prístup k internetu. Počet webových domén vo svete rastie a spolu s nimi aj počet ich užívateľov. Dosah internetu je obrovský, veľmi rýchly a ťažko kontrolovateľný. Vytvára tak „priestor“ ľuďom, ktorí sa napríklad snažia o zvrhávajúce internetových debát, narušenie vecnej diskusie alebo ovplyvňovanie mienky spoločnosti. S tým je spojené šírenie falošných správ a taktiež využívanie internetových trolov. Nie všetci užívatelia internetu sú dostatočne obozretní v selekcii informácií, ku ktorým sa vďaka internetu dostanú, resp. nevedia tieto informácie dostatočne rozoznať. Dochádza tak ku šíreniu správ, ktoré trolovia využívajú vo svoj prospech a snažia sa zvrhnúť diskusiu, podnecovať tak zníženie jej vecnosti, ba až šírenie nenávisti medzi diskutujúcimi. Aktuálnosť témy, identifikácie falošných správ a internetových trolov, je veľmi vysoká a čoraz viac vnímaná, keďže trolovia a falošné správy začínajú mať značnú moc nad atmosférou v online priestore, pričom ich vplyv na spoločnosť môže byť veľmi zreteľný.

V internetovom slangu je trol osoba, ktorá začína hádky, snaží sa uraziť ľudí, rozptýliť a zasiať rozkol uverejňovaním poburujúcich, odbočujúcich, alebo nesúvisiacich príspevkov v online konverzácií, vo fórach, komunitách a podobne. Úmyslom trola je vyprovokovať ostatných používateľov k emotívnym reakciám a odbočeniu od témy konverzácie, či už za účelom vlastného obohatenia, alebo len pobavenia. Internetový trol je osoba, ktorá sa často skrýva za anonymnú identitu.

V tejto práci sú opísané typy trolov, na ktorých môžeme najčastejšie natrafiť v online priestore, ako ich dokážeme rozpoznať, ako aj spôsoby, ktorými ich možno pomocou zberu údajov či strojového učenia identifikovať. Cieľom danej práce je navrhnúť prístup k budovaniu detekčného modelu umožňujúceho jednoduchšiu identifikáciu trolov figurujúcich vo „vodách“ internetu a jeho webových domén.

II. TROLIZMUS

Trolizmus alebo troling je pomenovanie istej formy anti-sociálneho správania sa na internete. Pomenovanie „Troll“ je prevzaté zo severskej mytológie. Boli to vysoké tvory, ktoré žili na okraji ľudských spoločenstiev, najmä v lesoch. V literatúre boli pomerne často vykresľované ako hlúpe,

škaredé a silné, pričom robili ľuďom zle. Skrývali sa pod mostami a strašili deti, alebo ich dokonca jedli.

S trolizmom sa najčastejšie stretávame na sociálnej sieti Facebook. Daný fakt môže byť zapríčinený tým, že táto sieť má najväčšiu základňu užívateľov. Podľa štúdie z roku 2013, ktorej sa zúčastnilo 2000 adolescentov vo veku 14 až 18 rokov, až tretina z nich sa podieľala na trolistickom obsahu v posledných šiestich mesiacoch, a jedna desatina sa priznala k tomu, že v minulosti aktívne tvorila alebo vytvorila trolistický obsah. Medzi vysokoškolskými študentmi sa vďaka prieskumu dopracovali k záveru, že trolmi sú častejšie užívatelia, ktorí zdieľajú obsah frekventovanejšie a až 74% zúčastnených na prieskume vyslovilo, že za posledný týždeň sa stali svedkami, alebo obeťou trola. Trolmi sa, taktiež, približne 5-krát častejšie medzi študentmi stávali muži ako ženy [1].

Počítačové hry zvyknú byť často útočiskom trolov. Správanie sa trolov počas hrania sa počítačových hier, sa mierne líši od trolov, ktorých nájdeme v bežnom prostredí internetu. Tak ako pri komunikácii v diskusiách, aj v hrách, s možnosťou textovej alebo hlasovej komunikácie, dochádza k trolizmu, a to formou slov. S trolizmom, v textovej forme, sa na sociálnom webe stretlo 97% opýtaných hráčov. Pri samotnom hraní hier sa však, častejšie stretli s trolizmom vo forme hovoreného slova (84% voči 72% pri textovej forme). S inými formami trolizmu vo videohrách sa stretlo iba 17% opýtaných. V týchto prípadoch najčastejšie išlo o hranie proti cieľu hry alebo znepríjemňovanie hrania svojim spoluhráčom (ako príklad možno uviesť ich blokovanie, zabíjanie, úmyselné strieľanie vlastných gólov a pod.). Do značnej miery sa s trolizmom medzi hernou komunitou stretávame vo forme sexizmu a rasizmu. Hráčky uviedli, že boli často atakované kvôli ich pohlaviu. Niektoré preto často v hrách, v ktorých je možné si zvoliť pohlavie avatara, volili to mužské, aby takto znížili počet poznámok na ich adresu. Najčastejšie, ako dôvod svojho trolizmu hráči uvádzali nudu alebo frustráciu [2].

A. Typy trolov

Z vyššie uvedených faktov možno usúdiť, že nie každý troll je rovnaký, nemá totožný cieľ a motiváciu, alebo spôsob, ktorým sa prezentuje, pričom forma komunikácie a prejavu sa môže líšiť. Rozpoznávame niekoľko základných typov trollov [3]:

- *Trol provokatér* – snaží sa vyprovokovať negatívnu reakciu. Spôsob komunikácie môže využívať často veľmi jednoduchý, napríklad pod tragickú správu napíše krátky ale poburujúci komentár (ako „haha“).
- *Trol cvičiteľ (Grooming troll)* – v súkromnej komunikácii pomocou správ sa snaží pomaly meniť zmyslenie prijímateľa k budúcemu jednoduchšiemu ovplyvňovaniu
- *Trol oponent* – sa snaží každému za každú cenu oponovať a vyvolať tak negatívnu reakciu. Často má vysokú kadenciu zverejňovania príspevkov. Typický je nekonzistenciou svojich názorov, bez náznakov o jej snahu, často sa profiluje do opozície voči väčšine diskutujúcich.
- *Sociálny troll* – tento typ trolla sa často stáva prostriedkom k šíreniu propagandy a využíva sa k zmene politickej, alebo spoločenskej situácie, najväčšie zastúpenie má v trollích farmách, kde jedna fyzická osoba môže predstavovať niekoľko takýchto trollov
- *Trol predavač* – využíva dezinformácie a polopravdy k propagácii pochybných služieb, alebo produktov. Jeho cieľom je vylákať peniaze podvodným spôsobom.
- *Trol urážač* – iba uráža diskutujúcich. Jeho cieľom je oberať diskutujúcich o energiu alebo vyvolať v nich rovnakú reakciu, aby tak demoralizoval ich správanie tým, že klesnú na jeho úroveň, a taktiež znížil takto hodnotu ich tvrdení.
- *Satirický troll* – využíva dobre vymyslenú satiru, ktorú je ťažké rozpoznať. Takáto sofistikovaná forma satiry sa aj ťažko vytvára, a preto je tento typ trollizmu menej efektívny, ako ostatné.
- *Trol vandal* – zvyčajne upravuje fakty v online článkoch a prekrúca ich tak, aby vyzneli vtipne, pričom niekedy celé odstavce vymazáva a nahrádza ich urážkami na komunitu, ktorej sa môže článok venovať
- *Trol odborník* - zvyčajne je v téme veľmi dobre zorientovaný, využíva poznatky z relevantných zdrojov, avšak, upravené tak, aby jemu vyhovovali.

B. Existujúce prístupy

Na internete už funguje niekoľko praktických nástrojov, ktoré nám s tým môžu pomôcť. Takisto sa v posledných rokoch týmto smerom upriamuje viacero výskumov s úspešnými výsledkami.

Haters News je online webový nástroj, ktorý pomáha s určením počtu nenávislivých príspevkov na sociálnych sieťach. Zameraný je na Reddit, Twitter a portál Hacker News. Jeho cieľom je

zistiť, v koľkých percentách svojich príspevkov, resp. statusov, sa užívateľ venoval hejtu, teda, uverejnil nenávisťný obsah. Spracováva posledných 500 zverejnených príspevkov daného užívateľa. Výstup je v grafickej podobe. Tá ma tvar podobný periodickej tabuľke prvkov, v podobe heat mapy, ktorá naznačuje mieru negatívosti jednotlivých príspevkov. Využíva k tomu vektorizáciu textu a logistickú regresiu. Jednotlivé okienka tabuľky sú taktiež označené začiatočnými písmenami príspevku.

Blbec.online je primárne nástroj, ktorý agreguje údaje o stránkach na facebooku, ich postoch a interakciách s nimi na svojej doméne. Jeho účelom je boj proti dezinformáciám, hoaxom a nenávisťi širenej prostredníctvom internetu. Jeden z nástrojov, ktorý doména blbec.online taktiež ponúka, je zameraný na správcov stránok na Facebooku. Tým môže pomôcť s vyhľadávaním nenávisťných komentárov na ich stránkach a automaticky ich skrývať. Vďaka tomuto nástroju vedia zamedziť šíreniu komentárov a príspevkov od nežiaducich profilov a stránok na stránkach pod svojou správou. Stačí sa registrovať pod svojím Facebook účtom, nastaviť niekoľko povolení. Potom stačí upraviť zoznam profilov, od ktorých nechceme, aby uverejňovali spam pod našimi príspevkami. Takto dochádza k zníženiu neetických príspevkov alebo falošných správ, či spamov.

ClueBot je automatizovaný užívateľ Wikipédie, teda bot, slúžiaci na automatickú detekciu vandalizmu wiki článkov. Prvá verzia ClueBota bola nasadená v roku 2007, odvtedy prešiel niekoľkými úpravami a verziami. K reverzácii vandalizovaných článkov využíva niekoľko metód strojového učenia. Prvá verzia bola založená na jednoduchých heuristikách a blackliste zakázaných slov. Súčasný model využíva niekoľko rôznych Naïve Bayes klasifikátorov, ktoré využívajú blacklisty, ale aj whitelisty slov. ClueBot dokáže úspešne odhaliť až 90% poškodzujúcich zmien v článkoch. Za minútu dokáže ClueBot overiť viac ako 9000 editácií.

Vedecké prístupy často využívajú Twitter a Reddit, pre ich rozsiahle API. Pomocou nich vedia získať údaje aj o zdieľaných príspevkoch, a ich ďalšom šírení, dáta o používaných hashtagoch, sledovaných a sledujúcich užívateľov a iné. Vďaka údajom unikátnosti URL adries v postoch, frekvencie postovania príspevkov za čas, podobnosti hashtagov dokážu pomocou metód strojového učenia pomerne presne identifikovať trollie účty využívané k spamovaniu. Vychádzajú pri tom z hypotézy, že spamboty využívajú často rovnaký kontext, zdieľajú podobný obsah a využívajú URL adresy rovnakej domény [4].

Identifikáciou trolov na Twitteri sa taktiež zaoberala štúdia z roku 2015, ktorá bola aplikovaná na reálne dáta kyberšikany na tejto sieti. Vychádzali pri tom z predpokladu, že každý profil trola je sledovaný (má folow) od skutočnej fyzickej osoby za týmto profilom. Tento predpoklad bol založený na fakte, že každý správca troliho profilu chce ostať informovaný o stave falošného profilu a diani na ňom, aj keď pod ním práve nie je prihlásený. Využívali pri tom dolovanie dát z textu a snažili sa nájsť podobnosti v písomnom prejave profilov, ktoré sa vzájomne sledovali [5].

III. NAVRHNUTÝ PRÍSTUP

Nami navrhnutý prístup bol implementovaný v nasledovných krokoch:

- Vytvorenie datasetu zo sociálnej siete Facebook použitím služby Export Comments na sťahovanie príspevkov súvisiacich s pandémiou SARS-Cov2 na Slovensku obsahujúci 2500 komentárov
- Z týchto komentárov bola manuálne selektovaná reprezentatívna množina so zachovaním pomeru troľích a netroľích komentárov
- Dataset bol anotovaný pomocou nasledovných tried:
 - Pozitívny a pozitívny vecný komentár
 - Neutrálny vecný komentár
 - Neutrálny komentár
 - Negatívny vecný komentár
 - Negatívny komentár
 - Provokatívny komentár
 - Silno provokatívny až nenávisťný komentár
- Extrahovanie základných vlastností textu
 - Celkový počet znakov príspevku
 - Počet slov príspevku
 - Priemerná dĺžka slova
 - Počet veľkých písmen v texte
 - Počet číslíc v texte
- Výber sady atribútov (knížnica *SelectPercentile* zo Scikit-learn pre výber atribútov textu s najvyšším skóre)
- Trénovanie modelov
 - Metóda podporných vektorov (SVM – Support Vectom Machine)

- Naivný Bayes klasifikátor (NB)
- Logistická regresia (LR)
- Vektorizácia textu (testy sa uskutočnili na dátach, ktoré boli aj na dátach, ktoré neboli pred-spracované pomocou TF-IDF)
- Testovanie modelov na rozpoznávanie trolov

Výsledky testov modelov naučených na dátach bez vektorizácie textu ako aj s vektorizáciou textu použitím troch metód strojového učenia a to SVM, NB a LR sú uvedené v Tabuľke 1. Najlepšie výsledky SVM sme dostali použitím RBF(radial basis function) kernel funkcie.

Tabuľka 1
Výsledky testov modelov natrénovaných metódami SVM, NB a LR s vektorizáciou textu aj bez nej.

SVM	Bez vektorizácie			Vektorizácia textu		
	Presnosť	Návratnosť	F1	Presnosť	Návratnosť	F1
Netroll	56%	100%	71%	43%	40%	41%
Troll	100%	8%	8%	36%	38%	37%
MNB	Presnosť	Návratnosť	F1	Presnosť	Návratnosť	F1
Netroll	89%	53%	67%	88%	47%	61%
Troll	63%	92%	75%	60%	92%	73%
LR	Presnosť	Návratnosť	F1	Presnosť	Návratnosť	F1
Netroll	90%	69%	78%	75%	67%	70%
Troll	67%	89%	76%	67%	77%	71%

Vektorizácia textu neprinesla očakávané zlepšenie výsledkov. Presnejšie výsledky pomocou dolovania dát z textu, by sa dali dosiahnuť aj pomocou použitia iného prístupu. Pri vektorizácii textu TD-IDF sa nepozerala na kontext slov vo vete ani na ich význam, ale iba na ich výskyt vo vetách, obsiahnutých v dokumente. Tieto slová často nadobúdajú rôznu podobu ich spájaním, skracovaním a ohýbaním. S týmto nám môže pomôcť stemming (redukcia odvodených slov), lematizácia (identifikácia významu slova vo vete) a tokenizácia. Pri týchto postupoch sa často využívajú slovníky. Angličtina je pomerne jednoduchý jazyk pre aplikovanie takýchto postupov. Slovenčina je na druhej strane náročnejší jazyk na spracovanie. V dnešnej dobe už existujú nástroje na báze neurónových sietí, ktoré ponúkajú aj predtrénované modely identifikácie významu slov vo vetách. Tieto modely sú založené na reálnych dátach a ponúkajú širokú paletu jazykov. Použitie daných metód je síce náročné na výpočtový výkon a veľkosť trénovacej množiny, avšak mohli by nám v budúcnosti ponúknuť lepšiu alternatívu.

IV. ZÁVER

V článku je uvedený popis prístupu k identifikácii trolov v online komunitách extrakciou základných vlastností textu a na takto získaných dátach aplikovanie metód strojového učenia. Ztroch použitých metód najlepšie výsledky dosiahol detekčný model naučený pomocou logistickej regresie. Je viacero možností, ako by sa dala práca rozšíriť, zaujímavý by bol najmä prístup s využitím dolovania v textových dátach pomocou neurónových sietí. Vylepšenie by mohlo nastať taktiež v zmene prístupu k získavaniu dát, pomocou iných aplikácií alebo vytvorenia vlastného riešenia. Pri nájdení vhodnej kombinácie prístupov, by bolo možné vytvoriť aplikáciu, napríklad vo forme rozšírenia webových prehliadačov, ktorá by sama dokázala v reálnom čase jednoducho trola identifikovať.

POĎAKOVANIE

Práca prezentovaná v tomto článku bola podporená Slovenskou agentúrou na podporu vedy a výskumu v rámci projektu VEGA 1/0685/21.

LITERATÚRA

- [1] F. HONGA, F., CHENGB, K. Correlation between university students'online trolling behaviorand online trolling victimization forms, current conditions, andpersonality traits; *Telematics and Informatics*; 35(1), 2018, pp.397-397.
- [2] TACKER, S., GRIFFITHS, M. D. An Exploratory Study of Trolling in Online Video Gaming. *International Journal of Cyber Behavior, Psychology and Learning*, 2(4), 2012, 17-33.
- [3] BERGHEL, H., BERLEAND, D. The Online Trolling Ecosystem. *Computer*, 51(8), 2018, pp. 44-51.
- [4] FAZIL, M., ABULAIISH, M. A Hybrid Approach for Detecting Automated Spammers in Twitter. *IEEE Transactions on Information Forensics and Security*, 13(11), 2018, pp. 2707-2719.
- [5] GALÁN-GARCIA, P., PUERTA, J. G. D. L., GÓMEZ, C. L., SANTOS, I., BRINGAS, P. G. Supervised machine learning for the detection of troll profiles in twitter social network: application to a real case of cyberbullying. *Logic Journal of IGPL*, 2015, jzv048.

Predikčné finančné modely časových radov s využitím vlnkových transformácií

¹*Tímea BALHOVÁ*, ²*Richard LONŠČÁK*, ³*Anna JADLOVSKÁ*

^{1,2,3}Department of Cybernetics and Artificial Intelligence, Faculty of Electrical Engineering and Informatics Technical University of Košice, Slovak Republic

¹timea.balhova@student.tuke.sk, ²richard.lonscak@tuke.sk, ³anna.jadlovska@tuke.sk

Abstrakt – Témou tohto príspevku je využitie aplikačných toolboxov programového prostredia MATLAB pre získanie predikčných finančných modelov. Článok vznikol na základe vlastnej diplomovej práce [1]. V jadre príspevku sú popísané základne poznatky o finančných časových radoch a o modeloch, ktoré sú využívané pri Box-Jenkinsonovej metóde, sú tu popísané vlnkové transformácie, ktoré sa používajú na predspracovanie dát pri tvorbe predikčných finančných modelov ako napríklad Haarova vlnková transformácia. Posledná časť príspevku je venovaná aplikácii teoretických poznatkov a simulačnému overeniu algoritmov pre vytvorenie predikčných modelov s využitím vlnkových transformácií.

Príušové slová – časové rady, ARMA model, ARIMA model, predikcia, vlnky, vlnková transformácia, Haarova transformácia, Daubechien4, MODWT, programové prostredie MATLAB

I. ÚVOD

Téma predikcie časových radov je stále aktuálna, stále existuje množstvo metód, ktoré je možné skúmať. Pri predikcii je dôležité, aby vytvorené predikcie boli čo najpresnejšie. Pre získanie presnejších výsledkov sa začala používať pri predspracovaní dát vlnková transformácia. Nakoľko časové rady často bývajú deterministické, zvykne sa podľa Wadiho a kol. [2] pri nich používať MODWT (maximum overlap discrete wavelet transform) – maximálne prekrytie diskkrétnej vlnkovej transformácie. Pomocou MODWT sa má vyriešiť obmedzenie DWT, v ktorom sa vyžaduje $T = 2^k$, kde k predstavuje celé kladné číslo. V roku 2011 Zhu a kol. využili pri predikcii časových radov model MODWT-ARMA, empirický režim rozkladu (EMD) a taktiež DWT. Model MODWT-ARMA je možné aplikovať na všetky typy dát [3]. Vo svojej štúdiu uviedli princíp tvorby modelu MODWT-ARMA, ktorý aplikovali na meteorologické dáta. Tento princíp pozostáva z dvoch krokov: v prvom sa pomocou vlnkového rozkladu časový rad rozdelí na detailnejšie časové rady, kde v druhom kroku na každý takýto detailnejší rad je aplikovaný predikčný model ARMA.

Okrem modelu MODWT-ARMA sa na modelovanie nestacionárnych časových radov využíva model MODWT-ARIMA [2][4]. Wadi a kol. [5] vo svojej štúdiu overovali na finančných dátach presnosť dekompozície pomocou dynamickej MODWT, Haarovej vlnkovej transformácie (DHWT) a Daubechienovej vlnkovej transformácie (dWT) na ktoré aplikovali ARIMA model. Najvšeobecnejšie porovnanie predikcie časových radov bolo vykonané v roku 2017 Zhangom a kol. [6] na meteorologických dátach. Pri výslednom porovnaní presnosti okrem modelov založených na vlnkových rozkladoch pre modely ARMA a ARIMA porovnávali aj základné modely ARMA a ARIMA bez použitia vlnkového rozkladu. Ako sa dalo predpokladať, najvyššiu presnosť vykazoval model ARIMA založený na vlnkovej transformácii. V roku 2018 Farima a Utami [7] vydali článok, v ktorom sa zamerali na predikciu výmenného kurzu amerického dolára voči indonézskej rupii. Vo svojej štúdiu využili MODWT s modelom ARMA.

II. MODELÝ ČASOVÝCH RADOV

Medzi najdôležitejšie kvantitatívne metódy analýzy ekonomických dát patrí práve analýza časových radov. Za hlavný cieľ analýzy časových radov sa považuje vytvorenie vhodného modelu na základe historických dát, ktoré nám slúžia na predikciu období v budúcnosti. Podľa Cipru [8] si môžeme predstaviť, že znalosť modelu zodpovedá znalosti algoritmu, podľa ktorého sa dáta

generujú počítačom. Časový rad je vo všeobecnosti náhodný proces, ktorý sledujeme v rôznom čase $t, t-1, t-2, \dots, t-p$. Písmeno p reprezentuje rad modelu, v tomto konkrétnom prípade to je rad autoregresného modelu – AR modelu. Jednotlivé hodnoty, ktoré budem sledovať v čase som si zadefinovala ako $y_t, y_{t-1}, y_{t-2}, \dots, y_{t-p}$.

Box-Jenkinsova metóda kladie hlavný dôraz na reziduálnu zložku, ktorá pozostáva z náhodných hodnôt, ktoré sú navzájom závislé [8]. Táto metóda obsahuje dva jednoduché modely, kde ten najjednoduchší je model kľzavých súčtov prvého rádu a ten druhý je autoregresný model. Stretáme sa tu aj so zmiešanými modelmi ARMA a integrovaným modelom ARIMA. Model ARMA (p, q) je definovaný nasledovne. V definícii ϕ_1 predstavuje autoregresný operátor a θ_1 je parameter procesu kľzavých súčtov.

$$y_t = \phi_1 \cdot y_{t-1} + \dots + \phi_p \cdot y_{t-p} + \epsilon_t + \theta_1 \cdot \epsilon_{t-1} + \dots + \theta_q \cdot \epsilon_{t-q} \quad (1)$$

Model ARIMA (p, d, q) je zovšeobecnený ARMA model, ktorý dokáže spracovať aj nestacionárne dáta. Počet diferencií udáva písmeno “ d ”. Tento model je definovaný pomocou operátora spätného posunu B .

$$\phi(B)w_t = \theta(B)\epsilon_t$$

$$w_t = \Delta^d y_t$$

Vo všeobecnosti je teda možné tento model zapísať ako

$$\phi(B)(1-B)^d y_t = \theta(B)\epsilon_t .$$

III. VLNKOVÉ TRANSFORMÁCIE NA PREDSPRACOVANIE DÁT

Vlnková transformácia predstavuje matematický model, ktorý transformuje pôvodný signál na inú doménu určenú na analýzu. Konkrétne sa tento signál transformuje na vlnky. Na začiatku vlnkového rozkladu máme pôvodný signál, ktorému prispôbíme materskú vlnku v najmenšej miere. Materská vlnka je základná funkcia a má svoj začiatok a aj koniec. Toto prispôbenie vytvára to, čo sa nazýva “fluktuácia” prvej vlnky a zvyšok sa nazýva “trend”. Potom sa zdvojnásobí časový rozsah materskej vlnky a prispôbíme ju prvej aproximácii. Toto vytvorí druhý detail vlnky a zvyšok je potom druhá aproximácia. A tak proces pokračuje, kým sa materská vlnka nerozšíri do takej miery, že pokryje celý rozsah signálu [9].

A. Diskrétna Haarova vlnková transformácia (DHWT)

Pri Haarovej transformácii je dôležité aj dodržať hodnotu T , počet vstupných údajov. Táto hodnota má byť taká, aby platilo $T = 2^k$. Základný princíp Haarovej transformácie je podľa Walkera [10] nasledovný.

Majme daný počet hodnôt T , kde T je párne číslo

$$y = (y_1, y_2, \dots, y_t).$$

V prvom kroku si vypočítame párový priemer čísel, ktorý následne pre-násobíme $\sqrt{2}$. Vieme to zapísať ako $s_t = [(y_{2t-1} + y_{2t})/2] * \sqrt{2}$ alebo nasledovne

$$s_t = \left(\frac{y_{2t-1} + y_{2t}}{\sqrt{2}} \right), \quad (2)$$

pre $t = 1, 2, 3, \dots, T/2$. Ďalej je potrebné si vypočítať fluktuáciu, čo vlastne predstavuje výkyvy signálu. Hodnoty fluktuácie získame párovým rozdielom predeleným dvomi a pre-násobeným $\sqrt{2}$. Vieme to zapísať ako $d_t = [(y_{2t-1} - y_{2t})/2] * \sqrt{2}$ alebo nasledovne

$$d_t = \left(\frac{y_{2t-1} - y_{2t}}{\sqrt{2}} \right), \quad (3)$$

pre $t = 1, 2, 3, \dots, T/2$. Zobrazíme si hodnoty pre prvú úroveň výpočtov Haarovej transformácie nasledovne

$$y \rightarrow (s^1 | d^1).$$

Pre ďalšiu úroveň transformácie vypočítame hodnoty rekurzívne podľa (2) s^2 a podľa (3) d^2 . Druhá úroveň transformácie bude mať teda nasledujúce zobrazenie:

$$y \rightarrow (s^2 | d^2 | d^1).$$

Následne je potrebné vykonať inverznú Haarovu transformáciu. Napríklad pre získanie vstupných hodnôt y_t z hodnôt $(s^1|d^1)$ to vykonáme nasledovne.

$$y = \left(\frac{s_1 + d_1}{\sqrt{2}}, \frac{s_1 - d_1}{\sqrt{2}}, \dots, \frac{s_{T/2} + d_{T/2}}{\sqrt{2}}, \frac{s_{T/2} - d_{T/2}}{\sqrt{2}} \right) \quad (4)$$

B. Daubechien4 (dWT)

Rozdiel medzi Daubechien4 a Haarovou transformáciou spočíva v spôsobe, akým sú definované funkcie pre výpočet škálovacích koeficientov a vlnkových koeficientov. Pri tejto transformácii sa v každom kroku transformácie sa na vstupné dáta aplikuje škálovací a aj vlnkový filter [10].

Koeficienty škálovacej funkcie α_l sú definované nasledovne:

$$\alpha_1 = \frac{1 + \sqrt{3}}{4\sqrt{2}}, \alpha_2 = \frac{3 + \sqrt{3}}{4\sqrt{2}}, \alpha_3 = \frac{3 - \sqrt{3}}{4\sqrt{2}}, \alpha_4 = \frac{1 - \sqrt{3}}{4\sqrt{2}}. \quad (5)$$

Signály škálovania D4 druhej úrovne sa vytvárajú opakovaním operácie, ktoré boli použité na prirodzenom základe $V_1^0, V_2^0, \dots, V_T^0$ na generovanie signálov škálovania prvej úrovne. Prvá úroveň má teda nasledujúci predpis

$$V_t^1 = \alpha_1 \cdot V_{2t-1}^0 + \alpha_2 \cdot V_{2t}^0 + \alpha_3 \cdot V_{2t+1}^0 + \alpha_4 \cdot V_{2t+2}^0 \quad (6)$$

Koeficienty pre vlnkovú funkciu β_l získame nasledovne:

$$\beta_1 = \frac{1 - \sqrt{3}}{4\sqrt{2}}, \beta_2 = \frac{\sqrt{3} - 3}{4\sqrt{2}}, \beta_3 = \frac{3 + \sqrt{3}}{4\sqrt{2}}, \beta_4 = \frac{-1 - \sqrt{3}}{4\sqrt{2}}. \quad (7)$$

Koeficienty vlnkovej funkcie sú tiež v každej iterácii sú posunuté o dve pozície. Prvá úroveň D4 pre vlnkový signál je daná nasledovne

$$W_t^1 = \beta_1 \cdot V_{2t-1}^0 + \beta_2 \cdot V_{2t}^0 + \beta_3 \cdot V_{2t+1}^0 + \beta_4 \cdot V_{2t+2}^0 \quad (8)$$

C. Maximálne prekrytie diskretnej vlnkovej transformácie (MODWT)

Medzi hlavnú výhodu MODWT oproti DWT patrí to, že dokáže spracovať akúkoľvek veľkosť vzorky. Ďalšou výhodou je to, že umožňuje rýchlejšie porovnanie medzi časovým radom a jeho rozkladom. Je to vlastne transformácia, ktorá porovnáva vlnkové koeficienty v každom časovom intervale s pôvodným signálom [7].

Pri tejto metóde sa najprv vypočítajú filtre pre zvolenú diskretnú vlnkovú transformáciu a následne je potrebné si vypočítať škálovací $s_{j,l}$ (9) a vlnkový filter $d_{j,l}$ (10) pre MODWT, ktoré sú definované nasledovne [7].

$$\tilde{s}_{1,l} = \frac{s_l}{2^{l/2}} \quad (9)$$

$$\tilde{d}_{1,l} = \frac{d_l}{2^{l/2}} \quad (10)$$

kde l predstavuje dĺžku filtra, v tomto prípade $l = 1, 2, \dots, L$ a j predstavuje level dekompozície. Pri využití D4 je $j=4$. Pre časový rad y_t s ľubovoľnou veľkosťou vzorky T sú vlnkové koeficienty W_j (11) a škálovacie koeficienty V_j (12) j -tej úrovne definované nasledovne:

$$W_j = \sum_{l=0}^{L_j-1} \tilde{d}_l \cdot y_{(t-l) \bmod T} \quad (11)$$

$$V_j = \sum_{l=0}^{L_j-1} \tilde{s}_l \cdot y_{(t-l) \bmod T} \quad (12)$$

kde $L_j = (2^j - 1)(L - 1) + 1$. Následne je definovaný vlnkový koeficient j -tej úrovne MODWT ako

$$W_{j,t} = \sum_{l=0}^{L_j-1} \tilde{d}_{j,l} \cdot y_{t-l \bmod T} \quad (13)$$

a škálovací koeficient ako

$$V_{j,t} = \sum_{l=0}^{L_j-1} \tilde{s}_{j,l} \cdot y_{t-l \bmod T} \quad (14)$$

Časový rad sa potom získa ako

$$y_t = \sum_{j=0}^{J_0} W_j^T W_j + V_{j_0}^T V_{j_0} = \sum_{j=0}^{J_0} D_j + S_{J_0}$$

IV. METÓDA TVORBY PREDIKČNÉHO MODELU

Dáta, s ktorými som pracovala predstavujú historické dáta Európskej centrálnej banky, ktoré sú uložené ako CSV súbor. Tieto dáta sú z devízového finančného trhu Slovenskej republiky a sú datované od roku 1999 až po rok 2022. Tento príspevok je zameraný na predikciu výmenného kurzu amerického dolára (USD). Počet vstupných dát bol $T = 256$, čo predstavuje 52 týždňov. Tieto dáta som si následne rozdelila na tréningové $T = 230$ a testovacie $T = 26$.

Následne som si vypočítala vlnkové transformácie pre 4 levely pomocou týchto transformácií: Haar, Daubechien4 a MODWT. Po vykonaní týchto vlnkových transformácií som identifikovala modely ARMA a ARIMA pre jednotlivé získané časti signálov pomocou aukorelačnej funkcie. Po identifikácii som vykonala odhad parametrov týchto modelov. Výsledný predikčný model predstavoval sumár všetkých vytvorených modelov s odhadnutými parametrami.

V ďalšom kroku som vykonala predikciu pomocou vytvorených modelov pre jednotlivé vlnkové transformácie. Na záver som pomocou výslednej odchýlky RMSE (15) vykonala porovnanie získaných modelov.

$$\sqrt{\sum_{t=1}^T \frac{(y_t - \hat{y}_t)^2}{T}} = \sqrt{\sum_{t=1}^T \frac{e_t^2}{T}} \quad (15)$$

V. APLIKÁCIA TVORBY PREDIKČNÉHO MODELU PRE ZVOLENÚ VLNKOVÚ TRANSFORMÁCIU

A. Vlnková transformácia pomocou Daubechien4

Na tréningových dátach som vykonala vlnkovú transformáciu pomocou Daubechien4 a získala som 4 čast', ktoré predstavujú vlnkové koeficienty a jednu časť, ktorá predstavuje aproximáciu.

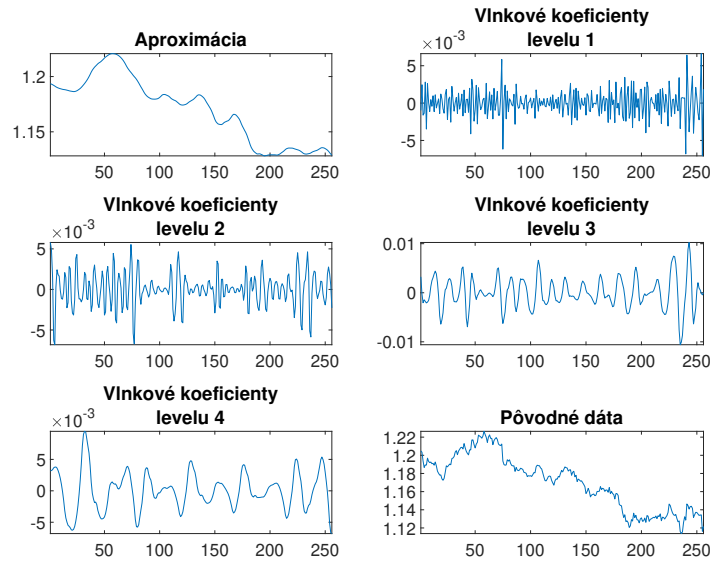


Fig. 1: Rekonštrukcia signálov dát po transformácii Daubechien4

B. Odhad parametrov transformovaného modelu

V ďalšom kroku som identifikovala modely pre jednotlivé časti signálov (1) získané po transformácii.

Následne som pre tieto modely odhadla parametre. Odhadnuté parametre pre prvý level fluktuácie pre model ARMA(3,1) sú

$$y_t = -0.80153 \cdot y_{t-1} - 0.6299 \cdot y_{t-2} - 0.2426 \cdot y_{t-3} + \epsilon_t - 0.95028 \cdot \epsilon_{t-1} \quad (16)$$

Table 1: Identifikované modely pre jednotlivé časti signálu s využitím D4

Časť signálu	Identifikovaný model
fluktuácia $d1$	ARMA(3,1)
fluktuácia $d1$	ARMA(3,1)
fluktuácia $d3$	ARMA(1,2)
fluktuácia $d4$	ARIMA(2,1,1)
aproximácia $a5$	ARIMA(1,2,1)

parametre pre druhý level fluktuácie pre model ARIMA(2,1,1) sú

$$y_t = 0.13113 \cdot y_{t-1} - 0.54589 \cdot y_{t-2} + \epsilon_t + 0.87434 \cdot \epsilon_{t-1} \quad (17)$$

parametre pre tretí level fluktuácie pre model ARMA(1,2) sú

$$y_t = 0.94431 \cdot y_{t-1} + \epsilon_t + 1.9258 \cdot \epsilon_{t-1} + 0.9258 \cdot \epsilon_{t-2} \quad (18)$$

parametre pre štvrtý level fluktuácie pre model ARIMA(2,1,1) sú

$$y_t = 1.2634 \cdot y_{t-1} - 0.44438 \cdot y_{t-2} + \epsilon_t + 0.1496 \cdot \epsilon_{t-1} \quad (19)$$

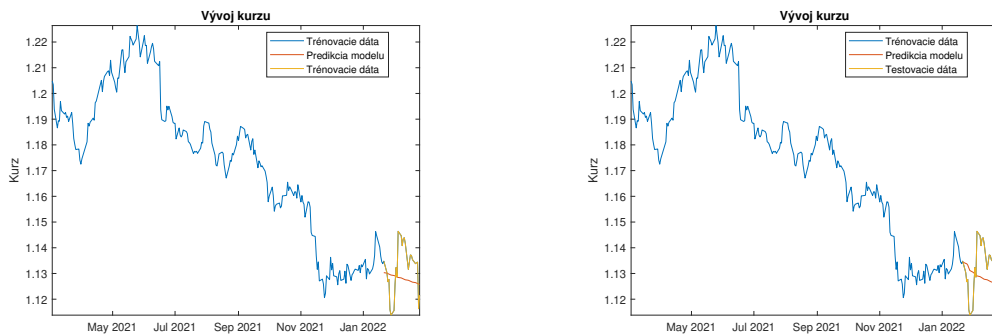
a parametre pre aproximáciu pre model ARIMA(1,2,1) sú

$$y_t = 0.3841 \cdot y_{t-1} + \epsilon_t + 0.18168 \cdot \epsilon_{t-1} \quad (20)$$

Výsledný predikčný model nám predstavuje sumár všetkých modelov (16 - 20).

VI. SIMULAČNÉ OVERENIE VYTVORENÝCH MODELOV

V tejto časti som si najskôr vytvorila predikcie a následne som ich porovnávala s pôvodnými dátami. V jednotlivých grafoch môžeme vidieť pôvodné hodnoty, ktoré boli pred predikciou, čiže tréningové dáta, testovacie dáta a výsledné predikcie po aplikácii vlnkových transformácií. Na grafoch môžeme vidieť, že vo všetkých štyroch grafoch majú predikčné hodnoty získané z vytvorených modelov klesajúci charakter [1].

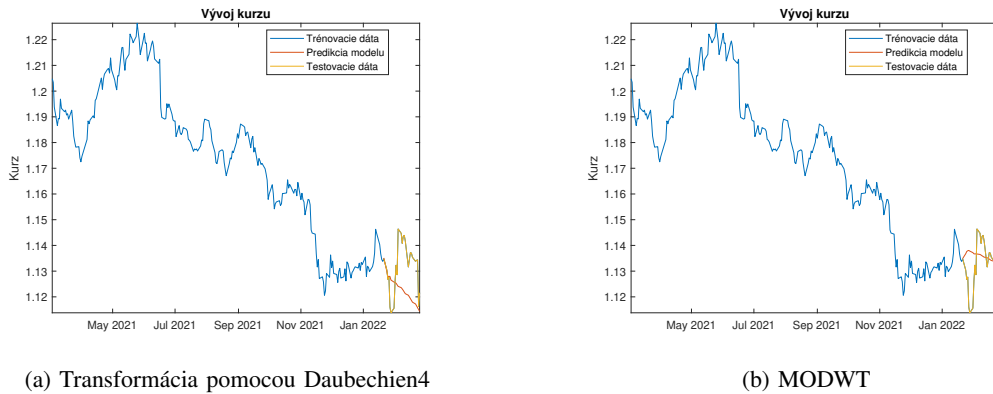


(a) Vlastná funkcia

(b) Využitá funkcia z MATLAB-u

Fig. 2: Predikcia vývoja kurzu pomocou Haarovej vlnkovej transformácie

Obrázok (2a) predstavuje predikciu, kde predikované hodnoty získané z vytvoreného modelu pomocou vlastnej funkcie majú klesajúci charakter, ktorý sa približuje k pôvodným hodnotám, až na výkyvy. Oproti modelu získanom pomocou funkcie z MATLAB-u (2b), má táto predikcia strmější spád. Obrázok (3a) zobrazuje predikované dáta, ktoré na začiatku kopírujú pôvodné hodnoty, no po čase nadobúdajú strmý spád. Na obrázku (3b) majú predikované hodnoty získané z vytvoreného modelu po transformácii MODWT oproti predchádzajúcim transformáciám najmenej strmý spád [1].



(a) Transformácia pomocou Daubechien4

(b) MODWT

Fig. 3: Predikcia vývoja kurzu pomocou ďalších vlnkových transformácií

Table 2: Porovnanie odchýlok modelov

Vytvorená predikcia	MSE	RMSE
Predikcia s využitím DHWT	1.2530×10^{-5}	0.0035
Predikcia s využitím D4	2.9053×10^{-6}	0.0017
Predikcia s využitím MODWT	2.3952×10^{-5}	0.0049
Predikcia s využitím vlastnej funkcie DHWT	1.0152×10^{-5}	0.0032
Predikcia s dátami bez vlnkovej transformácie	1.4358×10^{-5}	0.0038

V tabuľke sú uvedené výsledné odchýlky pre predikčné modely, ktoré boli vytvorené po aplikácii vlnkových transformácií a aj odchýlka modelu bez využitia týchto transformácií. Najpresnejšie výsledky boli dosiahnuté pomocou modelu, ktorý na transformáciu dát využil vlnkovú transformáciu pomocou Daubechien4.

VII. ZÁVER

Z výsledného porovnania odchýlok (2) si môžeme všimnúť, že najmenšiu odchýlku dosahoval model, s využitím vlnkovej transformácie pomocou Daubechien4 a naopak najvyššiu odchýlku dosahoval model s využitím MODWT, ktorý využil Haarove filtre. Dokonca aj predikčný model bez využitia vlnkovej transformácie mal menšiu odchýlku ako MODWT. Pri MODWT boli použité aj filtre D4, no pri daných vstupných dátach nebolo možné vytvoriť vhodné modely. Predikcia s využitím Haarovej vlnkovej transformácie (DHWT) pomocou vlastného vytvoreného algoritmu dosahovala nižšiu odchýlku ako DHWT pomocou funkcie, ktorú MATLAB obsahuje [1].

REFERENCIE

- [1] T. Balhová, *NÁVRH PREDIKČNÝCH FINANČNÝCH MODELOV S VYUŽITÍM APLIKAČNÝCH TOOLBOXOV PROGRAMOVÉHO PROSTREDIA MATLAB [diplomová práca]*. Košice: Technická univerzita, 2022.
- [2] S. A. Wadi, A. Hamarsheh, and H. Alwadi, "Maximum overlapping discrete wavelet transform in forecasting banking sector," *Applied Mathematical Sciences*, vol. 7, pp. 77–80, Apríl 2013.
- [3] L. Zhu, Y. Wang, and Q. Fan, "MODWT-ARMA model for time series prediction," *Applied Mathematical Modelling*, vol. 38, pp. 1859–1965, 2014.
- [4] T. Kriechbaumer, A. Angus, D. Parsons, and M. R. Casado, "An improved wavelet-ARIMA approach for forecasting metal prices," *Resour Policy*, vol. 39, pp. 32–41, 2014.
- [5] S. N. Sudrimo, K. Sadik, and I. M. Sumertajaya, "Analysis of Time Series Data Using Maximal Overlap Discrete Wavelet Transform Autoregressive Moving Average," *Applied Mathematical Sciences*, 2020.
- [6] H. Zhang, S. Zhang, P. Wang, Y. Qin, and H. Wang, "Forecasting of particulate matter time series using wavelet analysis and wavelet-ARMA/ARIMA model in Taiyuan, China,," *Journal of the Air & Waste Management Association*, vol. 67, pp. 776–788, Apríl 2017.
- [7] V. Z. Farima and H. Utam, "Peramalan Nilai Tukar Dolar Amerika terhadap Indonesia dengan Model Maximal Overlap Discrete Wavelet Transform-Autoregressive Moving Average," *Statistika*, vol. 6, pp. 73–77, Máj 2018.
- [8] T. Cipra, *Analýza časových řad s aplikacemi v ekonomii*. SNTL, 1986.
- [9] D. Goring, Orthogonal wavelet decomposition. [Online]. Available: <http://www.tideman.co.nz/Salalah/OrthWaveDecomp.html>
- [10] J. S. Walker, *A Primer on Wavelets and Their Scientific Applications (2nd ed.)*. Chapman and Hall/CRC, 2008.

Softvérová realizácia zobrazovacej jednotky UWB senzorovej siete

¹Adam RAČKO, ²Jana FORTES

¹Katedra elektroniky a multimediálnych telekomunikácií, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach

²Katedra matematiky a teoretickej informatiky, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach

¹Adam.Racko@student.tuke.sk, ²Jana.Fortes@tuke.sk

Abstrakt – UWB (z angl. *Ultra-WideBand*) radarové systémy schopné rýchlej detekcie a lokalizácie osôb majú širokospektrálne uplatnenie. V prípade vojny, prírodných katastrof alebo asistovaného života seniorov dokážu zachraňovať životy. Predkladaný príspevok prezentuje návrh softvérovej realizácie 2D zobrazovacej jednotky pre senzorovú sieť na báze UWB radarov. Hlavným cieľom vývoja aplikácie bolo znížiť nároky na výpočtový systém a ušetriť priestor na disku. Opis vývoja webovej a systémovej aplikácie je doplnený testovaním navrhutej zobrazovacej jednotky. Testovanie potvrdilo schopnosť klientskej aplikácie načítať vstupné dáta z externých súborov, ale aj zo servera, a to v reálnom čase. Overená a následne preukázaná bola aj podpora najpoužívanejších operačných systémov: Windows a Mac OS.

Kľúčové slová – UWB senzorová sieť, moderné webové technológie, grafický dizajn, vývoj aplikácie

I. ÚVOD

V roku 2022 už zrejme nikto nespochybnuje fakt, že technológie nám každým dňom čoraz viac uľahčujú život, ba dokonca v niektorých situáciách život zachraňujú. Medzi život zjednodušujúce a život zachraňujúce technológie patria aj ultra širokopásmové (z angl. *ultra wideband*, UWB) radarové systémy [1]. Použitie UWB radarových systémov v súčasnosti nie je ničím neobvyklým, práve naopak, využívajú sa pomerne často a v rôznych sférach všedného života. Najčastejšie sú potrebné v kritických situáciách ohrozujúcich zdravotný stav človeka, pretože predstavujú inovatívny a efektívny spôsob detekcie a lokalizácie osôb nachádzajúcich sa za prekážkou. Za prekážkou sa osoby môžu ocitnúť v prípade vojnových konfliktov, prírodných katastrof, teroristických útokov alebo nebezpečných momentov spojených s rastúcou kriminalitou. Na zobrazovanie polohy lokalizovaných osôb je nutné vytvoriť rýchly, bezpečný, vizuálne atraktívny a zároveň jednoduchý softvér.

Predkladaný príspevok, ktorý približuje riešenie načrtnutej úlohy, vychádza z diplomovej práce A. Račka [2]. Tá si v rámci praktickej časti kládla za cieľ na základe architektúry klient-server navrhnuť a zaviesť softvérové prostriedky pre realizáciu 2D zobrazovacej jednotky pre senzorovú sieť na báze UWB radarov. Úlohou bolo vyvinuté riešenie realizovať ako klientsku aplikáciu na bežnom počítači prostredníctvom voľne dostupných vývojových nástrojov. Riešenie malo podporovať jednoduchú konfiguráciu topológie použitej senzorovej siete s UWB radarmi minimálne v konfigurácii s 1, 2, alebo 3 uzlami. Softvérová aplikácia mala umožňovať spracovanie zaznamenaných dát z externých súborov a tiež spracovanie dát v reálnom čase získané zo servera.

Teoretická analýza technických požiadaviek ukázala, že medzi webové rámce (frameworks) vhodne využiteľné pre danú aplikáciu patria *VueJS* a *ElectronJS* [3], [4], [5]. *VueJS* je progresívny, reaktívny framework, ktorý je primárne určený na tvorbu interaktívnych webových rozhraní a *SPA* (z angl. *Single Page Application*). Progresívnym je práve preto, lebo je schopný sa inkrementálne adaptovať. V reaktivite sa vo veľkej miere inšpiruje *Reactom*, deklaratívnym vykresľovaním, virtuálnym *DOM* (z angl. *Document Object Model*) a objektovo orientovanou reprezentáciou *HTML* dokumentu. Opísaným spôsobom je schopný vytvárať si výstupný obraz vo *Vue* pomocou stromu komponentov. Rovnako ako *React* aj *Vue* sa primárne zameriava na základnú knižnicu, no navyše ponúka jednoducho pripojiteľný manažment stavov (angl. *global state management*) a router, čo z neho robí plnohodnotný framework s radou výhod, ako sú optimalizačné úsilie, šablóny

alebo *JSX* (z angl. *JavaScript eXtensible Markup Language*) a oficiálna podpora rozšírení [6]. Druhým zvoleným frameworkom je *ElectronJS*. Ide o dynamický, open source framework, ktorý umožňuje používateľovi vytvárať systémové aplikácie s pomocou *HTML5*, *CSS* a *JavaScript*. Akákoľvek webová aplikácia, ktorá je napísaná pomocou spomenutých technológií (*HTML*, *CSS*, *JS*, *VueJS*) môže využívať *ElectronJS*. Jeho súborový systém patrí *node.js API* (z angl. *Application Programming Interface*) a funguje na *Linuxe*, *Mac OS X* a *Windows*. Využíva *NPM* (z angl. *Node Package Manager*) modul, ktorý sa používa pre *JavaScript*. Skladá sa z natívneho menu pre dialógy a upozornenia. Inštalátory systému *Windows* nepotrebujú žiadnu konfiguráciu. Zložitými a zdĺhavými časťami vytvárania systémovej aplikácie sú zjednodušenia tzv. balenia, inštalácie, aktualizácie a poskytovania podpory pre natívne ponuky. Obsahuje široké *API* pre komunikáciu s operačným systémom s podrobnou dokumentáciou. *ElectronJS* sa do značnej miery stará o všetky tieto základné kroky, aby sa používateľ mohol sústrediť na jadro svojej aplikácie. Pracuje tak, že vezme hlavný súbor definovaný v *package.json* a spustí ho. Ten vygeneruje aplikačné okno, ktoré obsahuje render webovej aplikácie a natívne grafické rozhranie [7].

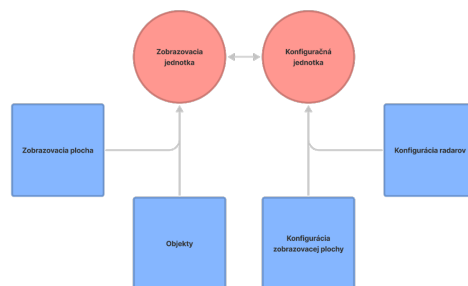
Vďaka vývojárovi Noah Klayman je možné každú *VueJS* webovú aplikáciu v krátkom čase pretvoriť na systémovú aplikáciu. Uvedený spôsob umožňuje používať aplikáciu bez potreby webového prehliadača na širokom spektre platforiem. Jeho projekt je voľne dostupný na *GitHube* a bol aktívne využitý aj v novovytvorenej softvérovej aplikácii [8].

Vychádzajúc z načrtnutých cieľov a technických prostriedkov má predkladaný príspevok nasledujúce členenie. Kapitola II vymedzuje funkcie požadované od navrhovanej aplikácie a zaoberá sa realizáciou jej vizuálnej stránky. Kapitola III popisuje štruktúru aplikácie, priebeh programovania nosných používateľských rozhraní a použitých knižníc, spracovanie údajov pred samotným zobrazením a testovaciu fázu vyvinutých funkcií a fungovania aplikácie práce. V závere príspevku sú zhrnuté vlastnosti vytvorenej softvérovej realizácie zobrazovacej jednotky UWB senzorovej siete a načrtnuté jej možné rozšírenia.

II. NÁVRH APLIKÁCIE

A. Požiadavky klienta

Hlavnou myšlienkou bolo navrhnuť aplikáciu, ktorá bude operovať ako samostatná jednotka bez potreby dodatočnej inštalácie ďalších programových doplnkov a knižníc. Na jej zostrojenie boli využité voľne dostupné programové prostriedky. Ideou bolo, aby jej výpočtový výkon bol dostatočný pre akúkoľvek platformu využitia. Spracovanie UWB radarových signálov, ktoré predchádza samotnému zobrazeniu výsledných pozícií sledovaných cieľov, bolo riešené v programovom prostredí *Matlab*. Samotné spracovanie sítě bolo správne, no jeho grafický dizajn a výpočtový výkon zobrazovania nebol príliš prijateľný. Je potrebné poznamenať, že *Matlab* obsahuje mnoho funkcií a zaberá veľké množstvo pamäte na disku, čo bolo dôvodom zostrojenia vlastnej zobrazovacej jednotky pre simuláciu v reálnom čase a tiež v off-line móde. Zobrazovacia jednotka by mala umožňovať grafickú vizualizáciu sledovaných pohybujúcich sa osôb v monitorovanej 2D oblasti a mala by byť nezávislá od existujúceho programového riešenia realizujúceho spracovanie UWB radarových signálov. Výraznou prednosťou aplikácie by malo byť atraktívne dizajnové prevedenie zobrazovacej jednotky a aj samotnej aplikácie. Aplikácia má obsahovať aj jednoduchú konfiguráciu radarov a zobrazovacej plochy. Kľúčovou funkcionalitou bude práve pohyb dizajnovo navrhnutých objektov po zobrazovacej ploche s informáciou o ich vzdialenosti od sledujúcich radarov, súradníc a pozorovacieho uhla. Zo zobrazovacej jednotky by malo byť evidentné, v akej polohe sa nachádzajú pozorované objekty, aký je počet pozorovaných objektov a aká je veľkosť pozorovacej plochy. Druhou funkcionalitou by malo byť prednastavenie pozorovanej plochy, súradníc, pozorovacieho uhla radarov a možnosti vybrať si koľko radarových systémov bude použitých pri danej simulácii.



Obr. 1 Návrh funkcionalít aplikácie.

Z Obr. 1 je možné vidieť, že nosným prvkom aplikácie budú dva komponenty, ktoré budú medzi sebou komunikovať a vykonávať funkcie im pridelené. Každý komponent bude obsahovať sadu funkcií samostatne vykonávajúcich konkrétnu funkcionálnu d'alej reprezentovanú v nosných komponentoch [2].

B. Dizajnové spracovanie

Pre navrhnutie grafického dizajnu je potrebný grafický dizajnový softvér, ktorý je určený na vytváranie vizuálneho obsahu alebo úpravu už existujúcich. Existuje mnoho grafických softvérov, takže dizajnér má možnosť voľby [9]. Pre vytvorenie dizajnu nami vyvíjanej aplikácie bol vybratý grafický softvér *Figma*, pretože je voľne dostupný a jednoduchý na používanie. *Figma* je nástroj na navrhovanie kolaboratívneho rozhrania, a na rozdiel od iných (napr. od *Sketchu*) funguje ako samostatná aplikácia založená v prehliadači, ale aj ako systémová aplikácia pre *Windows* a *MacOS*.



Obr. 2 Grafické riešenie pre konfiguračný a zobrazovací komponent.

Grafické riešenie pre konfiguračný a zobrazovací komponent je znázornené na Obr. 2. Úlohou konfiguračného komponentu je prijať dáta zadané užívateľom a odoslať ich zobrazovaciemu komponentu. Z tohto dôvodu bol navrhnutý ako formulár pre vyplňanie informácií, respektíve upravovania. V komponente sa nachádza viacero funkcií, akými sú napríklad načítavanie dát zo súboru pomocou vstupného tlačidla a vpísanie adresy pre pripojenie na server. Zobrazovací komponent znázorňuje samotnú zobrazovaciu plochu, objekty a informácie o ich x-ových a y-ových súradniciach, ale aj vzdialenosti od radarovej siete a pozorovaného uhla. Pri tejto editácii bolo potrebné uvažovať aj o nasledujúcej animácii objektov, preto boli navrhnuté tak, aby nevznikol problém pri finálnom testovaní simulácií sledovaných osôb. Problém, ktorý by mohol nastať je prekrytie daných objektov, a to by mohlo zapríčiniť nečitateľnosť informácií alebo nesprávne rozlíšenie.

III. VÝVOJ SYSTÉMOVEJ APLIKÁCIE

A. Štruktúra aplikácie

Štruktúra aplikácie popisuje interakcie medzi funkciami a komponentmi v aplikácii. Vďaka nej je zabezpečené fungovanie viacerých operácií súčasne. Architektúra webovej alebo systémovej aplikácie by sa mala zaoberať nielen efektivitou, ale aj spoľahlivosťou, škálovateľnosťou, bezpečnosťou a robustnosťou. Preto by mala zahŕňať všetky čiastkové komponenty aj výmenu informácií pre celú aplikáciu [2].

Hlavným komponentom je *App.vue*, ktorý spája dva dcérske komponenty a vytvára z nich fungujúcu aplikáciu. Zároveň je jeho úlohou pripraviť fungujúce prostredie pre operáciu s aplikáciou na strane klienta. Obsahuje tzv. karty, v ktorých sú vložené rozhrania a poskytuje im možnosť pohybovať sa medzi nimi. Do jadra štruktúry aplikácie importuje všetky potrebné knižnice a balíčky. Funguje ako zobrazovacia jednotka pre celú aplikáciu a má možnosť operovať so všetkými komponentmi nachádzajúcimi sa v hierarchii pod ním.

Prvým dcérske komponentom je *Animation.vue*, ktorého hlavnou funkciou je komunikovať s príbuznými komponentami a animovať zmenu iterácie v zobrazovacej jednotke aplikácie. Jeho príbuznými komponentami sú *App.vue* a *Basicconf.vue*. *Animation.vue* dostáva dáta o polohe radarov a objektov vo forme súradníc a renderuje ich v zobrazovacej jednotke. Zobrazovacia jednotka je vyplnená mriežkou pre jasnejšie odhadnutie pozícií cieľov. Pri radaroch je využitý aj parameter uhol pozorovania, ktorý umožňuje nastaviť aplikáciu presne podľa daného scenára.

Druhým dcérske komponentom je *Basicconf.vue*, ktorého hlavnou úlohou je umožniť operátorovi konfigurovať aplikáciu pomocou konfiguračného rozhrania. V konfiguračnom rozhraní je možné zvoliť si počet radarových systémov použitých pre daný scenár a nastaviť ich polohu a uhol

v rámci zobrazovacej jednotky. Zároveň je možné prednastaviť uvažovanú veľkosť pozorovanej plochy, v ktorej sa merajú súradnice objektov. V poslednom rade je užívateľ schopný rozhodnúť o tom v akej forme bude prijímať dáta - zo súboru alebo zo servera. Pre tieto dve možnosti sú vytvorené funkcie, z ktorých jedna poskytuje možnosť prijať dáta z vlastného počítača a druhá možnosť vložiť IP adresu daného servera.

B. Grafické používateľské rozhranie

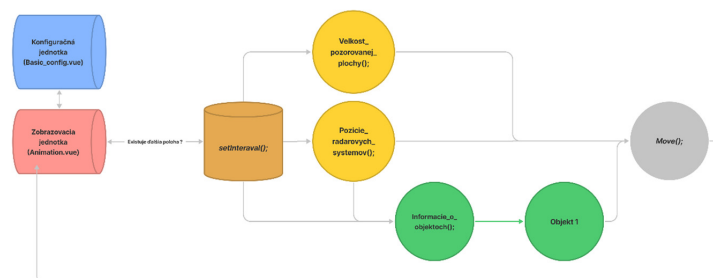
Grafické používateľské rozhranie tvorí komponent *Animation.vue*, ktorého hlavná úloha je opísaná v predchádzajúcej podkapitole. Na vytvorenie a vyvíjanie projektu bol použitý voľne dostupný zdrojový editor *Visual Studio Code*, ktorý obsahuje podporu pre *GitHub*, zvýraznenie syntaxe a podporu ladenia a refaktoriácie. Editor možno použiť s rôznymi programovacími jazykmi vrátane *HTML5*, *CSS3* a *JavaScript*. Zároveň podporuje moderné webové frameworky ako je *VueJS* a *ElectronJS*. Jeho hlavnou prednosťou je široké spektrum rozšírení, ktoré sú k dispozícii prostredníctvom centrálného úložiska.

Ďalšou inštaláciou bolo potrebné rozšíriť editor o *NPM* (z angl. *Node Package Manager*), vďaka ktorému môžeme prísť k inštalácii *VueJS* projektu. Inštalácia prebieha v príkazovom riadku editora. *CLI* (z angl. *Command Line Input*) je program, ktorý prijíma textový vstup na spustenie funkcií operačného systému. To dáva prístup k vytvoreniu adresára pre zamýšľaný projekt, v ktorom bude nainštalovaný *VueJS* balíček. Uvedený typ inštalácie zaručuje rýchle zostrojenie webovej aplikácie so všetkými potrebnými nástrojmi. Poskytuje nastavenia pre moderný frontend framework.

Vo vytvorenom projekte sa v tejto fáze nachádza iba komponent *App.vue*. Po následnom vytvorení dvoch dcérskych komponentov je potrebné zamyslieť sa, ako bude medzi nimi prebiehať komunikácia a ako bude vyriešený problém s animáciou objektov v grafickom používateľskom rozhraní. Komunikácia bude prebiehať prostredníctvom nástroja *Event Bus* v oboch komponentoch. Ide o výkonnú schopnosť, ktorá umožňuje komunikáciu mimo rodičovských komponentov. Všetky dáta potrebné na odoslanie sú najprv uložené do v-modelu, ktorý ich posiela do premenných a tie sa ďalej odosielajú do príslušného komponentu. Problém s animáciou bude vyriešený pomocou vopred nainštalovanej knižnice, vďaka ktorej je možné voľne pohybovať s objektmi po zvolenej ploche. Funkcia, ktorú knižnica obsahuje sa nazýva *Move.js*. Ide o ideálne riešenie, ktoré vytvára animácie *CSS3* pomocou jednoduchých funkcií. Poskytuje rozhranie *JavaScript API* na vytváranie daných animácií.

Problém však nastáva pri viacerých iteráciách, keďže nevieme koľko pohybov objekt vykoná pred ukončením jeho animácie. Riešením je vytvorenie slučky, ktorej známymi parametrami bude počet iterácií a rýchlosť animácie. Počet iterácií je možné zistiť zo súboru súradníc a rýchlosť animácie závisí od daného merania radarov. Pre vytvorenie slučky je najlepším riešením použiť známu funkciu *setInterval* umožňujúcu nastavenie rýchlosti opakovania animácie. Vnútro slučky musí obsahovať premennú, ktorá bude informovať o počte iterácií v danom momente. Funkcia *setInterval* obalí uje funkciu *move.js*, čím zabezpečuje opakovanie pohybu objektu podľa vstupných súradníc.

Po vyriešení problému s animáciou sa môžeme posunúť ďalej na problém vytvorenia objektu a zobrazovacej plochy. Vytváranie objektov a zobrazovacej plochy bude riešené pomocou *HTML* elementov, ktoré ponúkajú možnosť vyskladať si webovú aplikáciu s nespočetne veľa možnými elementmi. Okrem toho bude potrebné použiť *CSS*, ktoré dopomáha dané elementy upraviť podľa navrhnutého grafického dizajnu. Detaily tohto postupu sú opísané v [2]. Opísané implementované funkcie pracujú v postupnosti schématicky zobrazené na Obr. 3.



Obr. 3 Štruktúra *Animation.vue* komponentu.

C. Základná konfigurácia používateľského rozhrania

Základnú konfiguráciu používateľského rozhrania tvorí príslušný formulár umožňujúci zadávať užívateľom vstupné dáta a vue element *file-reader*, ktorého úlohou je načítať dáta zo súboru pre následné zobrazenie. Prvou fázou je implementovanie vhodného formulára. Formulár tvoria *HTML* elementy *input*, obalené tagom *form*. Pre konkrétny radarový systém sa v kóde nachádza osobitný tag *form* so vstupom pre zadávanie vstupných súradníc radarového systému. Zároveň je užívateľovi ponúknutá možnosť vypnúť alebo zapnúť zobrazenie systému. Táto funkcionálna bola pridaná, pretože existujú merania, ktoré pracujú iba s určitým počtom radarových systémov. V aplikácii je možné vybrať si jeden, dva alebo tri radarové systémy. Vypnutie a zapnutie definuje element *checkbox*. V aplikácii sa tiež nachádza vstup pre zadávanie hodnôt uhla pozorovania, aby bol operátor schopný nasmerovať radary podľa reálneho merania. Vue element *file-reader* sa nachádza na vrchu v grafickom dizajne, čo umožňuje operátorovi uvedomiť si, že prvým krokom je nahráť súbor so súradnicami z vlastného počítača. Návrh pridávania súboru je realizovaný pomocou jednoduchého kliknutia na tlačidlo, ktoré používateľ a presmeruje na adresár jeho počítača. Vstupné súradnice sa dajú vložiť do aplikácie aj pomocou servera. Takúto možnosť poskytuje textový formulár pre zadávanie IP adresy. Posledným pridaným formulárom je nastavenie veľkosti pozorovanej plochy. Ide o element *input*, ktorý je vopred vyplnený hodnotami 60 metrov na šírku a 40 metrov na výšku, čo definuje maximálnu veľkosť zobrazovanej plochy.

Rovnako ako pri prvom komponente, tak aj pri *Basicconf.vue* bola vytvorená slučka *setInterval*, ktorá zaručuje synchronizáciu dát medzi komponentmi. Na úplnom začiatku je priam nevyhnutné načítať zdrojové dáta pre zobrazovaciu jednotku. Okrem toho je dôležité do formulára vyplniť informácie o pozíciách radarových systémov a veľkosti pozorovanej plochy. Zobrazovacia plocha sa danému meraniu prispôbi až po úspešnom načítaní údajov a následne sa spustí simulácia dát. Pre lepšie pochopenie postupnosti využívania funkcií bol vytvorený Obr. 4, ktorý zobrazuje postupnosť implementovaných funkcií.



Obr. 4 Štruktúra *Basicconf.vue* komponentu.

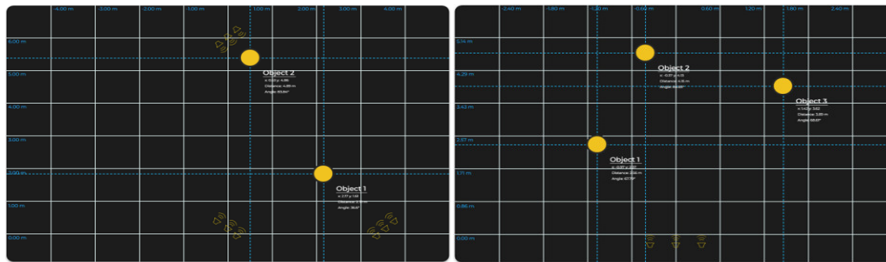
D. Spracovanie údajov pre zobrazenie

Spracovanie údajov pre následné zobrazenie v aplikácii prebieha v programovacom prostredí *Matlab*. Výstupný formát dát predstavuje trojrozmerná matica so súradnicami, ktorú je nevyhnutné pred zobrazením upraviť do tvaru *JSON* súboru. Pre tento konkrétny prípad bol vytvorený skript v *Matlabe* s názvom *saveAsJSON.m*, ktorý daný problém rieši. Skript vytvára šablónu *JSON* údajov a následne ich uloží do počítača v tvare *.json*. Takto štruktúrované dáta sú pripravené na vstup do vyvíjanej aplikácie pre zobrazenie objektu.

E. Testovanie

Testovanie aplikácie sa realizovalo manuálnou formou, pričom sa overila funkčnosť, použiteľnosť, rozhranie, kompatibilita a výkon aplikácie. Testovanie prebiehalo na vopred zvolených scenároch, pri ktorých boli využité merania s jedným, dvoma a tromi cieľmi. Súčasne bolo testované rozloženie radarových systémov a veľkosť pozorovacej plochy podľa reálnych meraní, pričom sa otestovalo rozloženie senzorovej siete pozostávajúcej z maximálne troch radarových systémov. Vstupné dáta do aplikácie boli načítané z počítača alebo posielané pomocou lokálne

vytvoreného JSON servera. Výstup zobrazovacej jednotky pre dva vybrané scenáre je ilustrovaný na Obr. 5. V prvom prípade išlo o off-line dáta z 3-uzlovej senzovej siete, ktorou boli sledované dve pohybujúce sa osoby, v druhom prípade išlo o online prijímané dáta zodpovedajúce sledovaniu troch ľudí pomocou jedného UWB radarového systému.



Obr. 5 Testovanie aplikácie na nameraných UWB radarových dátach.

Kompatibilita aplikácie bola overená aj na operačnom systéme *Mac OS*, čo znamená, že finálny produkt predstavuje aplikáciu vo formáte *.exe* a *.dmg*. Použitý výkon aplikácie predstavoval pri testovaní zhruba 40 MB pamäte RAM.

IV. ZÁVER

Predkladaný príspevok sa zaoberal problematikou vizualizácie zobrazenia lokalizovaných osôb v monitorovanom priestore pomocou moderných webových technológií. Hlavným cieľom práce bolo vyvinúť aplikačný softvér zabezpečujúci načítanie údajov z interných, ale aj externých úložísk, nakonfigurovať radarový systém a zobrazovaciu plochu podľa reálneho merania a následne zobraziť lokalizované osoby v navrhutej zobrazovacej jednotke. Vyvinutý softvér bol vytvorený pomocou moderných technológií ako je *VueJS* a *ElectronJS*, po ktorých je v súčasnosti vysoký dopyt veľkých či malých IT firiem. Dôležitou súčasťou práce je aj dizajnové spracovanie aplikácie navrhnuté v digitálnom dizajnovom prostredí *Figma*. Okrem samotného grafického a softvérového vývinu aplikácie bolo opísané aj testovanie aplikácie v dvoch operačných systémoch - *Windows* a *Mac OS*. K dispozícii je aj webová verzia softvérovej aplikácie.

Testovanie prebiehalo manuálnou formou na UWB radarových dátach poskytnutých Katedrou elektroniky a multimediálnych telekomunikácií TUKE. Dáta boli načítané z testovaného počítača (tzv. off-line zobrazenie) alebo posielané zo servera vytvoreného na testovanie (tzv. online zobrazenie). Keďže dáta, ktoré vstupujú do aplikácie, treba najprv upraviť na formát *JSON*, bol taktiež vytvorený skript v *Matlabe* zohľadňujúci uvedenú požiadavku. Potenciál pre budúci vývoj je napr. v mobilnej verzii takejto aplikácie, ktorej funkcionality by bola rovnaká, avšak na kompaktniejšom zariadení.

POĎAKOVANIE

Táto práca bola podporovaná Agentúrou na podporu výskumu a vývoja na základe projektu č. APVV-18-0373 a vedeckou grantovou agentúrou Ministerstva školstva, vedy, výskumu a športu SR a Slovenskej akadémie vied (VEGA) na základe projektu č. 1/0584/20.

LITERATÚRA

- [1] Kocur, D. et al., *Detekcia, lokalizácia a sledovanie osôb UWB radarmi krátkého dosahu*. Technická univerzita v Košiciach, Košice, 2016. ISBN 987-80-553-2648-1
- [2] Račko, A., *Softvérová realizácia zobrazovacej jednotky UWB senzovej siete*. Diplomová práca, Technická univerzita v Košiciach, Košice, 2022.
- [3] Study: Requirements in Web Applications: Examples & Challenges. [online]. Working Scholars. [citované 13.4.2022]. Dostupné na internete < <https://study.com/academy/lesson/requirements-in-web-applications-exampleschallenges.html> >.
- [4] GeeksforGeeks: Web Technology. [online]. GeeksforGeeks. [citované 14.4.2022]. Dostupné na internete < <https://www.geeksforgeeks.org/web-technology/> >.
- [5] Smetana, P., *Technológie pre webové služby a webové aplikácie*. Diplomová práca, Ekonomická univerzita v Bratislave, Bratislava, 2014.
- [6] Vue.js: The Progressive JavaScript Framework. [online]. Vue.js. [citované 14.4.2022]. Dostupné na internete < <https://vuejs.org/?msclkid=87400ff6bb1811ec9470c20644ae3d38> >.
- [7] Electron: Build cross-platform desktop apps with JavaScript, HTML, and CSS. [online]. Electron.js. [citované 14.4.2022]. Dostupné na internete < <https://www.electronjs.org/?msclkid=13e545c1bb1911ecbf4ce0b966e535f3> >.
- [8] Github: Nklayman/vue-cli-plugin-electron-builder, [online]. Github, Inc. [citované 14.4.2022]. Dostupné na internete < <https://github.com/nklayman/vue-cli-plugin-electronbuilder> >.
- [9] Rasmussen University: What is graphic design. [online]. Rasmussen College, LLC. [citované 14.4.2022]. Dostupné na internete < <https://www.rasmussen.edu/degrees/design/blog/what-is-graphic-design/> >.

Computer Security through Gamification

Anita TÓTHOVÁ, Anton BALÁŽ, Branislav MADOSŤ

Department of Computers and Informatics, Faculty of Electrical Engineering and Informatics Technical University of Košice, Slovak Republic

anita.tothova@student.tuke.sk, anton.balaz@tuke.sk, branislav.mados@tuke.sk

Abstract – This paper reviews a summary on the topic of Gamification in education. At the theoretical level, we focus on the meaning of Gamification and subsequently we introduce the concept of Internet Security. In the practical part, we create a system to support education through Gamification.

Keywords – education, game, gamification, learning, security

I. INTRODUCTION

The number of Internet users around the world is constantly growing and many are coming into contact with some form of cyber threat daily, which they are unable to deal with because they do not have enough information. This work aims to combine the idea of *Gamification* with education regarding Internet Security. The result is a system that allows users of all ages to test their knowledge of Security issues in a more easier way.

II. GAMIFICATION

First of all, we should define the term *Gamification*. The basic concept of *Gamification* is not new, but the word itself is a relatively new addition to the dictionary. *Gamification* is the use of gameplay elements in a non-gaming environment. It represents a technology that attempts to influence user behavior through elements of game design.[1] It is effective, mainly because it uses people's natural desire for competition and success. *Gamification* is also often used in applications and websites to motivate people to meet personal challenges, such as weight loss goals or language learning.

A. The MDA Framework

There are various design principles for game design to provide a positive user experience, such as integration and measuring progress, quick feedback, setting long-term and short-term goals, or being rewarded for completing assigned tasks. The MDA framework, i.e. mechanics, dynamics, and aesthetics, is important for the application of these basic principles.[2] The game must be interesting and bring meaningful value. Otherwise, our effort to attract the player's attention will not pay off.

Game mechanics describes the components of the game at the level of data presentation and algorithms. It has a great effect on user motivation. Common game mechanics include points used for rewards, scoreboards that offer users the opportunity to compare with other users in the system, or rewards by which users are compensated for success.[2]

Game dynamics describes the behavior of the mechanics during the run, so it is the reason for the user's motivational behavior towards the game mechanics.[2]

Game aesthetics means how the player uses and enjoys the game. How the game affects his senses. It is engaging and tempting to explore new environments, thus allowing self-realization.[2]

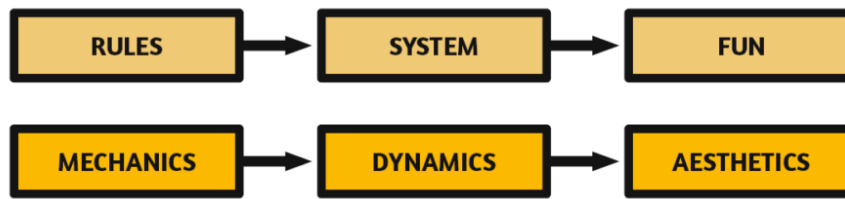


Fig. 1 MDA Framework[3]

B. Motivation

When discussing *Gamification*, it is essential to mention motivation. Motivation is generally divided into internal and external. External motivations are motivations that come from outside the individual and often include external rewards such as trophies, money, or social recognition. Internal motivations are motivations that come from within an individual, such as the feeling of satisfaction when solving a puzzle.[2]

Richard Bartle, game researcher, defined four types of players, also representing the different types of motivation that players may have. These types are the Killer, the Achiever, the Socializer, and the Explorer.[4]

Killers are highly competitive users who like to challenge other users and want to beat them, their key goal is a triumph. They want to be the best at the game and winning is what motivates them.[4]

Achievers' main stimulus is the collection of points, progression to a higher level, and higher evaluation. They want to show others how well they are doing in the game. They like to collect badges and display them.[4]

Socializers are the type of users who use the application as a bridge to connect and interact with other users. The community is a vital stimulus for users of this type. They like to work together to get bigger achievements than they can do on their own.[4]

Explorers are the type of users who want to learn about the application. Examining the application is therefore their main goal. They are not very interested in points or prices. The reward for them is a discovery itself.[4]

C. Game Design Elements

Game design elements are the building blocks of *Gamification* service packages. In shaping the game, elements of game design can support motivation to participate, stabilize user involvement and increase their creativity. Game elements include game points, which are allocated automatically for achieving predetermined goals, social points, assigned by other users, e.g. community ratings, redeemable points that represent currency in the game that can be spent on buying virtual or real goods, scoreboards that allow instant comparisons of users' performances, background stories to provide narration or virtual identity through game avatars.[2]

D. Influence of Gamification on Behaviour

One of the goals of *Gamification* is to influence user behavior. To influence behavior, we need to understand how behavior is created and what affects it.

Culture has many definitions and shapes people's reactions and preferences for computer systems and communication. Customs play an important role in culture because they facilitate decision-making in everyday life. In gamified applications/services, culture could also affect people's relationships. This means that consideration of behavioral and cultural patterns is crucial for effective use of *Gamification* in user behavior. [2]

Social factors can also influence behavior through social standards. For example, people and communities are often governed by standards through which members behave in a special way that affects their acceptance of opinions.[2]

A two-step procedure is required to influence behavior. First of all, it is necessary to create new behavior and then eliminate unwanted ones.

E. Game and Gamification

When discussing *Gamification* we should understand what a game is. The game has certain characteristics to be considered a game. It must have firm rules, sufficiently defined to avoid disputes. It must have variable and quantitatively measurable results, i.e. it must be adapted to the abilities and knowledge of the players, and the result of the game must always be designed so that no misunderstandings occur. And least, the devotion of the player to the result is needed.

Another sign of the game is that it has to be fun. In principle, we can distinguish 4 types of entertainment, namely easy fun, which is something that normally entertains us in games, for example, when something explodes, hard fun, i.e. challenges or solving tasks, people fun as socialization or teamwork, and last but not least, serious fun, when, for example, we solve a task that in the end, it will help someone.[5]

Gamified projects should have a cognitive, emotional, and social impact. Not everyone is willing to play or change their ways of learning, and there are two big obstacles to consider when creating gamified education. The first is that education is perceived by adults as a serious activity and *Gamification* is characterized by its light communication. The second is that a complicated game tends to cause confusion and frustration for participants. The game should have simple rules, details should be added with the gradual development of the game, and players need to emphasize the benefits of the game.

F. Gamification in Education

The main problems of modern education are related to the lack of commitment and motivation of students to actively participate in the learning process. For this reason, teachers use new techniques and approaches to provoke student activity and motivate them to participate. One possible solution is to reward the efforts and results achieved with an award, which leads to increased motivation to participate and the activity. E-learning creates favorable conditions for the implementation of *Gamification*. *Gamification* affects students' behavior, commitment, and motivation, which can lead to improved knowledge and skills.[6]

The development of an effective strategy for the implementation of *Gamification* in e-learning presupposes an in-depth analysis of current conditions and available software tools. The main steps of the strategy include the determination of student characteristics, the definition of learning objectives, the creation of educational content and activities for *Gamification* and the addition of game elements and mechanisms.[6]

III. INTRODUCTION TO INTERNET SECURITY

Information Security is defined as the protection of information and security systems against unauthorized access, use, disclosure, disruption, modification, or destruction. In general, security represents the goal to protect our assets from attackers, worms, power failure, vandalism, or another undesirable case. Defining when we can consider something safe is a challenge, but we can turn the question around, so defining when we are unsure is much easier. There we can include the use of weak passwords, downloading infected programs, or opening dangerous e-mail attachments.[7] The good thing is that if we can point out areas in the environment that can cause its uncertainty, we can take steps to alleviate these problems.

A. CIA Triad

When talking about Information Security, it is useful to have a framework that we can use as a basis. The basic concept of Information Security consists of confidentiality, integrity, and availability, also known as the CIA triad.[7]

Confidentiality is an essential part of privacy and relates to our ability to protect data from those who are not authorized to view it. Confidentiality can be compromised by the loss of a laptop containing data, in the case of a person looking over our shoulder while typing a password, or by sending an e-mail attachment to the wrong person. Integrity refers to the ability to prevent unauthorized or unwanted changes to our data. It may mean an unauthorized change or deletion of our data or parts thereof, or it may mean an authorized but unwanted change, as well as the deletion of our data. Availability refers to the ability to access our data when we need it. Loss of availability can involve a wide range of interruptions anywhere in the chain that allows us to access our data.[7]



Fig. 2 CIA Triad[8]

Several measures can be taken to help maintain confidentiality, including multi-factor authentication, strong passwords, encryption, or assigning users with appropriate levels of user rights. The two main ways to ensure confidentiality is cryptography and access control. To prevent unwanted modifications and to ensure that information can be restored in the event of a change, it is essential to perform regular backups, version checks, and access verification and efficiently manage access privileges. Integrity can also be ensured, for example, by using a hashing algorithm. Popular methods used to protect organizations from loss of availability include keeping all critical systems up to date, DDOS protection, redundancy, firewalls, proxy servers, and using access control. Availability ensures that authorized users of the system have timely and uninterrupted access to information in the system and the network.[9]

IV. SYSTEM PROTOTYPE PROPOSAL

A. Game Design

When designing the solution, it was necessary to pay attention to two key aspects, simple interpretation and the use of adequate means to ensure user motivation. The basis for achieving this goal was to create game scenarios that are sufficiently understandable and engaging at the same time.

The result of the solution design is a game concept that connects individual scenarios into a game loop. The user appears in the created game in the role of a candidate for a Cyber Security specialist. The game loop consists of a 30-day period, during which the user has the opportunity to test his or her knowledge in the field of Information Security. At the beginning of the game, the user has the opportunity to choose his avatar, with whom he can identify. Game mechanics used in the levels include quizzes, selection of suitable options, correct data sorting, setting secure passwords, finding differences, identifying associations, identifying fraudulent e-mails and websites, and encrypting messages.

One way to motivate the user is a reward mechanism. It consists in evaluating individual tasks, but also obtaining various hidden trophies. Trophies can be obtained by completing various side challenges. They are therefore an optional part of the game content and represent a form of reward that rewards the user for completing the secondary tasks. The rating comes after each level and is determined by the user's success. The method of evaluation is different for each type of game mechanics.

The overall evaluation of the game, based on which it is decided whether the candidate will remain in his position, is based on partial evaluations throughout the game. For success, it is necessary to obtain a partial rating above 65% for at least 20 days out of a total of 30.

To keep the user's attention, we chose a combination of different game mechanics when designing the game. These differ in the information they want to provide to the user and at the same time in the way, they are implemented.

1) *Quiz*: The quiz is a game mechanic in which the user is asked a question about cyber security and has a choice of 3 options. There is only one correct answer to each question. After choosing the correct answer, the user gets 100% in their evaluation. On the contrary, after choosing the wrong answer, the user is rated at 0%.

2) *Multiple Selection*: Multiple selection or selection of suitable options is a game mechanic in which the user has to choose several correct options. With this type of game mechanics, it is also possible to obtain a partial rating in case of an incomplete answer. There are a total of 5 answers, of which 3 are correct.

3) *Data Sorting*: Data Sorting is a game mechanic in which the user is shown several steps describing a certain operation, such as a symmetric encryption procedure, but in the wrong order. His task is to arrange the given steps correctly. When evaluating this game mechanic, the user receives a partial rating for each correctly arranged step.

4) *Setting a secure password*: Setting a secure password is the mechanic in which user input is expected to be entered. For 100% success, all the requirements for setting a strong password, which requires a lower-case and upper-case letter, a number, a special character, and its total length should be at least 8 characters, must be met.

5) *Finding differences*: Differences Finding is a game mechanic in which the user has to find objects that do not fit into the displayed scene in some way. The number of clicks is limited to prevent accidental findings. For each item found, the user is assigned a partial rating.

6) *Identification of associations*: Association determination is the mechanics of connecting entities belonging to each other. By clicking on two entities, an association represented by a line is created.

7) *Identification of fraudulent e-mails and websites*: Phishing identification is a game mechanic in which a user has access to email content or a preview of a website. The goal is to examine the content and evaluate whether this is an example of Phishing.

8) *Message encryption*: Message encryption is a mechanic where the user has access to the unencrypted e-mail content and the table on which the content is to be encrypted. After rewriting the original message into an encrypted form, the user's success is evaluated.

B. Implementation

Most games include continuous data storage. These are, for example, the game settings or the saved progress of the player. As part of the solution of this work, several methods of data storage were used. In the game menu, it is possible to set the volume in the game and the language used for the translation. To save these player preferences, `PlayerPrefs` is used, which is a part of the Unity Game Engine. `PlayerPrefs` also stores data of the current game, such as the current playing time or the name and selected avatar of the player.

Within the game cycle, the user's progress is saved automatically. The data is saved after clicking the button to return to the menu or within the evaluation of the level of success. Each time a new game is started, a new file is generated in which the progress of the game is saved.

The game cycle consists of scenes that represent the individual days of the trial period. Within each scene, a specific interpretation of one of the game mechanics is then prepared. The user has the option to reject the daily task.

V. EVALUATION

The solution of the work consisted of introducing the concept of *Gamification*, analyzing the current state of the problem, and finally creating a system to support security education through *Gamification*. At the core of the work, we prepared a solution proposal, which results in a fully functional, bilingual, computer game created using Unity Game Engine. It is intended primarily for companies and schools, but also for individuals who are interested in the topic of Internet Security. Using various game mechanics, it tests the user's knowledge in the field of Information Security. The player's external motivation is supported by continuous evaluation and rewards in the form of trophies. Internal motivation occurs when a player is satisfied with solving a task. The created game is easy to use with most mechanics allowing mouse control only. The prepared tasks test knowledge from various areas of Internet Security. The game contains a total of 30 prepared tasks, along with four side bonus tasks.

A. Survey Results

The final prototype of the game was tested by six users, who finally completed a questionnaire with 14 questions regarding their success, motivation, observed system deficiencies, and preferences. The respondents were in the age category from 18 to 35 years of age with secondary education and university education of the first and second degrees. The addressed respondents had basic knowledge in the field of Information Security.

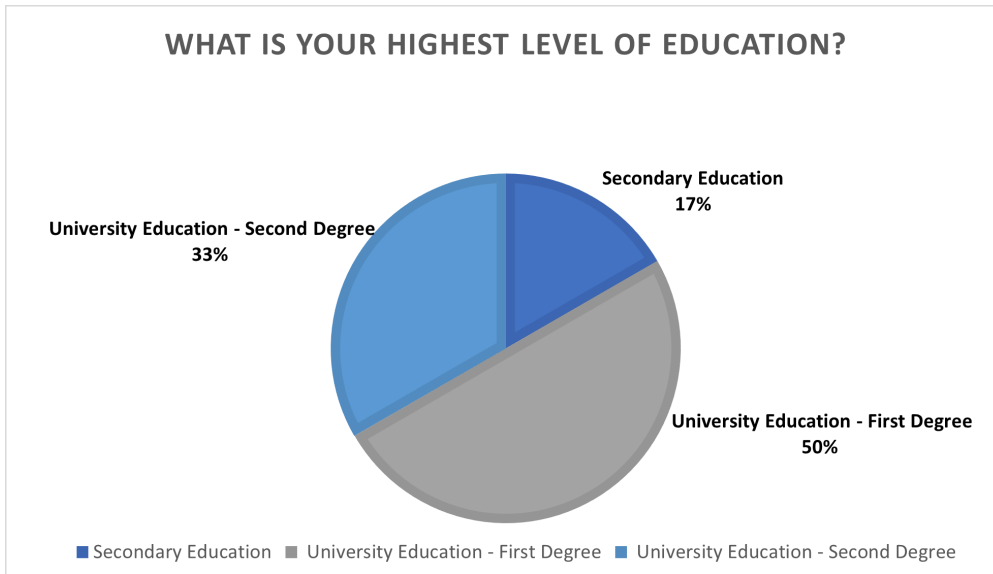


Fig. 3 Graph for the highest level of education of respondents

All respondents managed to finish the game, with the game time lasting from 19 to 26 minutes. The average overall user rating ranged from 66% to 88%. One-half of the users managed to get 4 and the other 2 out of a total of 5 trophies.

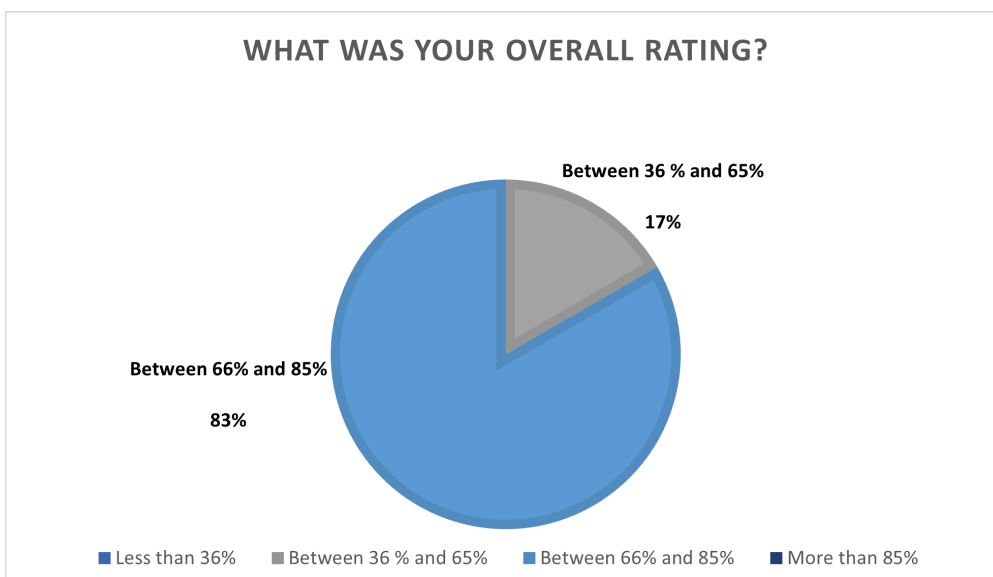


Fig. 4 Graph for the overall evaluation of respondents

A scale from 1 to 5 was used to assess the questions about the interface and game control, as well as the variability and difficulty of the tasks, where 1 represents full disagreement with the opinion stated in the question and 5 represents full agreement. When evaluating the appropriate difficulty of the tasks, up to five users equally chose the value 3, which indicates indecision in the given statement. The last user selected a value of 2, which means that he or she does not agree with the difficulty level setting.

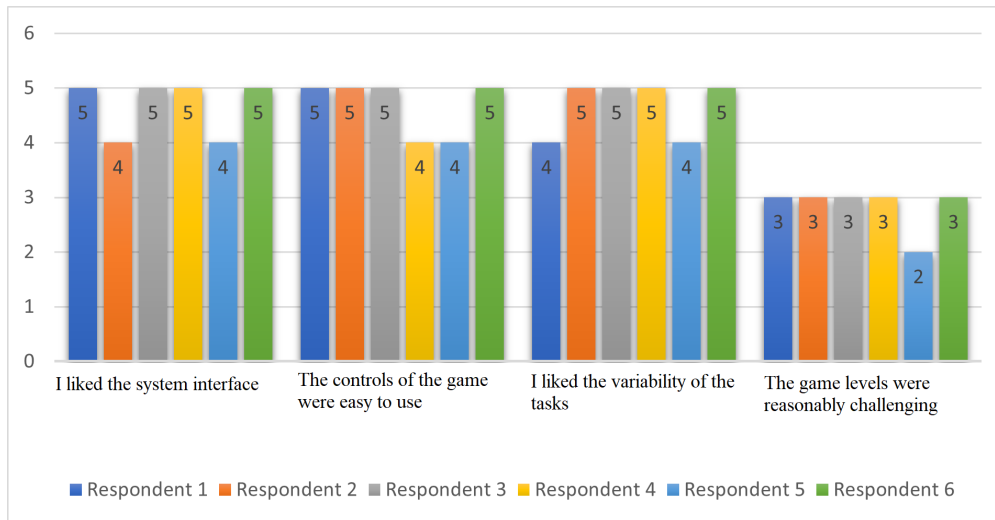


Fig. 5 Graph for scale evaluation of respondents

Based on the answers obtained, we can conclude that the primary motivation for completing the game was satisfaction from solving the task and trying to achieve the best possible rating. Positively evaluated elements of the game included the variability of challenges, the overall presentation of tasks, and the possibility of obtaining trophies. Shortcomings were noted, for example, low difficulty in the game mechanics of finding differences or the lack of additional information about the correctness of the solution. Popular gaming mechanisms included detecting phishing content and identifying associations. On the contrary, the least popular were message encryption and correct data sorting.

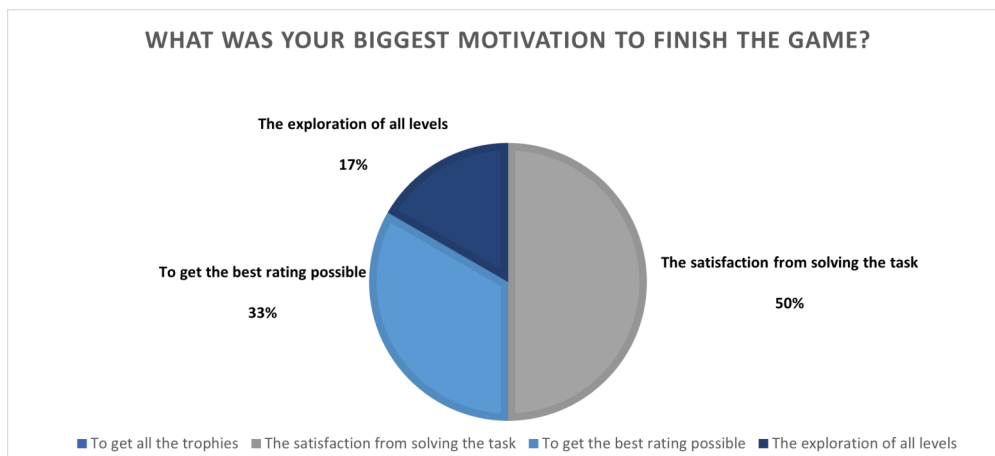


Fig. 6 Graph for respondents' motivation

By analyzing the collected answers, we can deduce that the prepared prototype meets the motivational component of *Gamification* thanks to its evaluation system and the variability of the prepared tasks. In terms of possible improvements, it is possible to optimize the complexity of individual tasks and adapt it to the target group. The acquired knowledge can be used for further expansion of the system.

VI. CONCLUSION

The aim of this work was the design and implementation of a system to support education in the field of Cyber Security using game elements. The result of the work is a game designed for personal computers implemented in the C# programming language using Unity Game Engine. The purpose of the created game is to provide the user with information related to Security in an entertaining way. For this purpose, several types of tasks have been prepared, which provide information on various areas of Information Security. The advantage of the system is the variety of game mechanics to support the user's motivation and at the same time test his attention in various ways. In the future, it would be possible to prepare this system in the form of a web

interface, which would ensure an easier expansion of the system among potential users. This also leads to a possible extension to the shared rating system, where users could compare their results. This would increase the external motivation to deliver better results and the overall motivation to continue playing. Possible extensions include the creation of an admin interface, within which it would be possible to define new tasks for pre-prepared task types. In this way, the system would offer the possibility of easily expandable content and at the same time the possibility of greater diversification.

ACKNOWLEDGEMENT

This work was supported by KEGA Agency of the Ministry of Education, Science, Research, and Sport of the Slovak Republic under Grant No. 002TUKE-4/2021 Implementation of Modern Methods and Education Forms in the Area of Cybersecurity towards Requirements of Labour Market.

REFERENCES

- [1] D. I. Blohm and P. D. J. M. Leimeister, "Gamification," *BISE – CATCHWORD*, 2013. [Online]. Available: https://doc.rero.ch/record/311550/files/12599_2013_Article_273.pdf
- [2] S. Stieglitz, C. Lattemann, S. Robra-Bissantz, R. Zarnekow, and T. Brockmann, Eds., *Gamification*. Springer, 2017. [Online]. Available: <https://EconPapers.repec.org/RePEc:spr:proiis:978-3-319-45557-0>
- [3] (2018) Mda in game development. [Online]. Available: http://ava-game.com/wp-content/uploads/2018/04/MDA_Framework-1024x342.png
- [4] (2021) Bartle's player types for gamification. [Online]. Available: <https://www.interaction-design.org/literature/article/bartle-s-player-types-for-gamification>
- [5] (2016) Základy gamifikácie: Čo treba vedieť skôr, ako začnete. [Online]. Available: <https://visibility.sk/blog/zaklady-gamifikacie-co-treba-vediet-skor-zacnete/>
- [6] L. Y. Gabriela Kiryakova, Nadezhda Angelova, "Gamification in education," *ACADEMIA*, 2014. [Online]. Available: <https://www.academia.edu/download/53993982/293-Kiryakova.pdf>
- [7] J. Andress, *The Basics of Information Security: Understanding the Fundamentals of InfoSec in Theory and Practice*, ser. Syngress basics series. Elsevier Science, 2014. [Online]. Available: <https://books.google.sk/books?id=9NI0AwAAQBAJ>
- [8] (2021) What are the 3 principles of information security? [Online]. Available: <https://images.squarespace-cdn.com/content/v1/5ef07334254666386ed45a91/1608262717294-MSU10IM5OR3JH523TK7Y/CIA+Triad.png?format=1000w>
- [9] (2018) The cia triad: The key to improving your information security[20/12/2021]. [Online]. Available: <https://commisum.com/blog-articles/the-cia-triad-the-key-to-improving-your-information-security/>

Pochopenie Vzniku Tsunami Pomocou Metód Dátovej Analytiky

Dominika CULBOVÁ, Viera ANDERKOVÁ, František BABIČ

Katedra kybernetiky a umelej inteligencie, Fakulta elektrotechniky a informatiky,
Technická univerzita v Košiciach, Letná 1/9, 040 01 Košice-Sever, Slovenská
republika

dominika.culbova@student.tuke.sk, viera.anderkova@tuke.sk, frantisek.babic@tuke.sk

Abstrakt — Tento článok popisuje aplikáciu vybraných metód prediktívneho a popisného dolovania s cieľom extrahovať nové a potenciálne zaujímavé znalosti z historických dát o výskyte tsunami. Tieto dáta pochádzali z databázy NGDC/WDS Global Tsunami Database. Analytický proces bol realizovaný v súlade s procesným modelom CRISP-DM. Celkovo sme vytvorili 131 modelov klasifikujúcich, či došlo k tsunami udalosti, alebo nie. Na základe výsledkov sme dospeli k záveru, že najvhodnejším modelom pre daný prípad je Rozhodovací strom, keďže dokázal najlepšie klasifikovať negatívnu triedu a súčasne dosiahol dobré výsledky pre vyhodnotenie situácie, kedy nastalo tsunami. Pomocou zhukovej analýzy sme vytvorili popisný model, na základe ktorého sme objavili viacero podobností, resp. odlišností, medzi udalosťami tsunami. Keďže účinky tsunami nie je možné zastaviť, potenciál našej práce je v oblasti zameranej na minimalizáciu ich následkov. Vzhľadom na formu možno povedať, že ide o podklady pre ľudí z danej sféry, ktorí by sprostredkované poznatky mohli využívať pri procese rozhodovania alebo v prípadných softvérových riešeniach.

Kľúčové slová — tsunami, klasifikácia, zhuková analýza

I. ÚVOD

Ľudstvo je už od svojho počiatku ohrozované rôznymi prírodnými katastrofami. Tsunami, séria extrémne dlhých vln spôsobená veľkým a náhlym premiestnením oceánu, patrí medzi najnebezpečnejšie a najsilnejšie z nich. Aj keď len mála časť zo vzniknutých udalostí spôsobí obrovské škody, ich následky majú tak tragický a devastačný rozmer, že ovplyvnia spoločnosť na veľmi dlhú dobu, prípadne už navždy. Medzi nich môžeme zaradiť straty na životoch, zranenia, zdevastované obydlia, finančné a materiálne škody, kontaminácia vodných zdrojov, únik nebezpečného materiálu do životného prostredia a mnoho ďalších.

Problémom je, že pôsobnosť tsunami nemožno zabrániť, preto je jediným riešením jej včasná predikcia. Vďaka nej je následne možné implementovať rôzne druhy patričných opatrení, ktoré dokážu minimalizovať potenciálne škody.

A. Súvisiace práce

Vzhľadom na náročnosť predikovania účinnosti tsunami, najmä z dôvodu veľkého množstva interferovaných faktorov, sa komunita vedcov neustále venuje ďalšiemu výskumu danej oblasti. Na základe toho sme uskutočnili rozsiahlu analýzu existujúcich štúdií zameraných na riešenie tohto problému. Súhrm relevantných poznatkov, primárne s ohľadom na metódy dátovej analytiky, ktoré nás ovplyvnili pri výbere nami použitých metód, zhrnieme v nasledujúcej časti.

V štúdiu [1] sa autori zamerali na predikciu stupňa nebezpečenstva tsunami prostredníctvom algoritmu C4.5, ktorý je založený na báze generovania rozhodovacieho stromu. Konkrétny výber odôvodnili jeho presnejšou klasifikáciou oproti metódam CHART a CHAID fungujúcim na rovnakom princípe. Vstupná dátová množina obsahovala 2 961 záznamov rozdelených na tréningovú a testovaciu množinu v rôznych pomeroch, pričom všetky údaje pochádzali z oblasti Pangandaran. Úspešnosť modelu bola vo všetkých prípadoch väčšia ako 97%. Aj keď navrhovaný model dosiahol veľmi kladné hodnotenie, je dôležité podotknúť, že pracoval s údajmi iba z jednej geografickej oblasti, čo nezaručuje jeho vysokú kvalitu aj v iných regiónoch sveta. Zároveň vyhodnocoval nebezpečenstvo len na základe dvoch parametrov, čo môže byť diskutabilné, keďže pôsobnosť tsunami je ovplyvnená viacerými faktormi.

V navrhovanom inteligentnom systéme Earthosys [2] je implementovaný algoritmus Random Forest s cieľom predpovedať vznik tsunami na základe informácií o zemetrasení. Autori riešenia zdôvodnili výber klasifikátora jeho výpočtovou rýchlosťou a vysokou mierou úspešnosti aj pri

nevyváženom dátovom súbore. Vstupnými parametrami modelu sú magnitúda, hĺbka ohniska a oblasť epicentra. Ak sa epicentrum nachádza na pevnine, zohľadňuje sa aj jeho vzdialenosť od najbližšieho pobrežného bodu. Na základe týchto údajov sa vykoná predpoveď, či ide o údaje spôsobujúce tsunami, alebo nie. V rámci štúdie bola použitá dátová množina z NOAA (National Oceanic and Atmospheric Administration), ktorú tvorilo 955 záznamov, z toho 25% (249) slúžilo na testovanie. Vyhodnotenie kvality modelu bolo realizované s ohľadom na obe klasifikované triedy prostredníctvom presnosti, návratnosti a F1 skóre. Výsledná, priemerná hodnota pre všetky metriky bola 0.98. Celková úspešnosť modelu vyšla na 97,99%, čo bolo potvrdené aj vykonaním krížovej validácie. Výsledky modelu boli veľmi pozitívne, avšak veľkosť použitej dátovej množiny ich mohla skresľovať.

Autori štúdie [3] navrhli systém zameraný na zisťovanie príchodu tsunami pomocou údajov o hladine mora, ako je výška vlny, tlak, rýchlosť vetra, príliv, odliv atď. V systéme implementovali algoritmus k-NN (k-Nearest Neighbors), kde $k = 10$, a ako metriku vzdialenosti definovali Euklidovskú vzdialenosť. Príklady, ktoré model vyhodnotil ako nezodpovedajúce stanovenej prahovej hodnote, boli následne vstupom do metódy k-Means. Tá bola aplikovaná pre automatické rozpoznávanie nových tried, aj keď model netrénoval s ich inštanciami. V situácii, kedy trieda vytvorená zhlukovaním vykazovala abnormálne hodnoty, bolo vydané varovanie pred vlnou tsunami. V opačnom prípade bola pridaná do tréningovej množiny pre budúcu klasifikáciu.

B. Použité metódy

V rámci analytického procesu sme postupovali v súlade s procesným modelom CRISP-DM [4], ktorý obsahuje šesť fáz: pochopenie cieľa, pochopenie dát, príprava dát, modelovanie, vyhodnotenie, nasadenie.

Rozhodovací strom [5] je jedným z najpoužívanejších klasifikátorov. Jeho súčasťou je konštrukcia stromovej štruktúry, ktorá vizuálne zobrazuje sériu postupných rozhodnutí vykonaných na dosiahnutie konkrétneho výsledku. Je zložená zo súboru uzlov obsahujúcich testovacie podmienky – interné uzly, a označenia tried – listové uzly. Počiatočný uzol stromu, z ktorého začína jeho vetvenie, sa nazýva *root* (koreň). Vyznačuje sa svojou jednoduchosťou a vysokou úrovňou interpretácie.

Random Forest [6] umožňuje dosiahnuť efektívnejšie výsledky ako Rozhodovací strom, keďže jeho princíp spočíva vo vytváraní niekoľkých menších rozhodovacích stromov pomocou náhodných vzoriek z tréningovej množiny. Tie sú následne spojené do jedného celku nazývaného *forest* (les). Výsledná predikcia vzniká z väčšinového hlasovania, resp. priemerných predpovedí jednotlivých rozhodovacích stromov.

K-NN [7] patrí do kategórie lenivého učenia, čo znamená, že žiadna zo znalostí nie je generovaná vopred. V prípade, že je zaslaná požiadavka na klasifikáciu nového príkladu, identifikuje sa jeho k-najbližších susedov pomocou funkcie vzdialenosti. Existuje viacero možností daného výpočtu, najčastejšie používaným je Euklidovská vzdialenosť. Testovaný príklad je klasifikovaný na základe identifikovanej triedy.

SVM [8] spočíva v nájdení nadroviny, ktorá rozdeľuje množinu údajov do dvoch skupín bodov, pričom každá z nich reprezentuje jednu z klasifikovaných tried. Optimálna nad rovina je tá, ktorá dosahuje maximálnu možnú vzdialenosť od akéhokoľvek bodu v rámci tréningovej množiny. Body nachádzajúce sa najbližšie k nej sa nazývajú podporné vektory a možno ich považovať za kritické prvky súboru údajov, keďže ich odstránenie z množiny by spôsobilo zmenu polohy deliacej nadroviny.

Naive Bayes [9] je založený na Bayesovej vete definujúcej výpočet podmienených pravdepodobností, čo je vlastne pravdepodobnosť jednej udalosti vzhľadom na výskyt inej. V rámci algoritmu sa tento výpočet vykoná pre každú triedu. Výsledkom predikcie je trieda s najväčšou hodnotou spomenutej pravdepodobnosti. Pri tomto procese je silný predpoklad nezávislosti jednotlivých atribútov.

Logistická regresia [10] klasifikuje prostredníctvom logistickej funkcie, ktorá sa používa na predpovedanie hodnôt pravdepodobnosti z intervalu $<0;1>$. Aby výsledkom algoritmu bola jedna z dvoch očakávaných tried, jeho súčasťou je hraničná hodnota, ktorá rozhoduje o výslednej klasifikovanej triede. Ak je pravdepodobnosť väčšia ako hraničná hodnota, priradí sa hodnota 1. V opačnom prípade je klasifikovaná hodnota 0.

K-Means [11] je podmienený vopred definovaným počtom zhlukov, ktorý je reprezentovaný v parametri k . Na zistenie jeho optimálneho počtu existuje niekoľko spôsobov. V našej práci sme aplikovali 3 z nich, a to Elbow metódu spočívajúcu vo výpočte súčtu štvorcových vzdialeností od každého bodu k jeho priradenému stredu, Silhouette Coefficient [12] počítajúci priemerný koeficient siluety pre všetky vzorky, Calinski-Harabasz Index, pri ktorom ide o výpočet rozptylu medzi klastrami a v rámci nich. Na základe stanoveného počtu k je do dátového priestoru prvkov

umiestené príslušné množstvo počiatkových jadier zhlukov. Po priradení každého dátového bodu k najbližšiemu jadru sa poloha jadier prepočítava. Tento krok sa opakuje dokým nie je splnené niektoré z kritérií zastavenia – žiadne dátové body nemenia zhluky, minimalizuje sa súčet vzdialeností, dosiahne sa maximálny počet iterácií.

II. ANALYTICKÝ PROCES

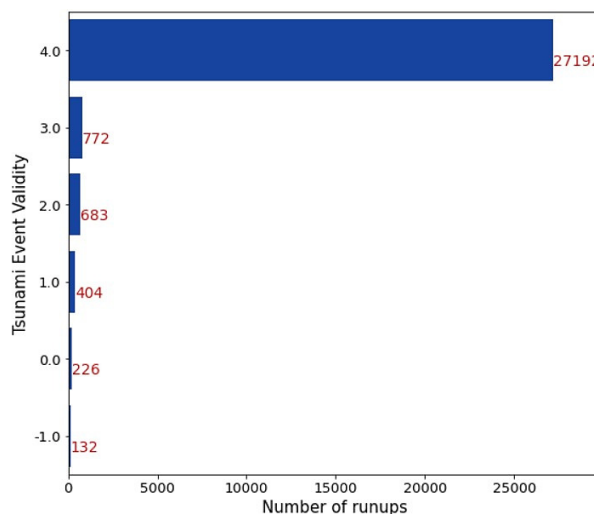
Táto kapitola popisuje realizované kroky a operácie v jednotlivých fázach procesného modelu CRISP-DM.

A. Pochopenie cieľa

Minimalizácia následkov tsunami udalostí, vzhľadom na ich závažnosť, je nanajvyš potrebná. V oblasti prediktívnej analýzy sa zameriame na vytvorenie modelov klasifikujúcich, či ide o účinky tsunami, alebo nie. Cieľom našej práce je teda klasifikovanie validity udalosti a nie predikcia jej budúceho vzniku. Prostredníctvom popisného modelu budeme skúmať existenciu podobností, resp. odlišností, medzi udalosťami.

B. Pochopenie dát

Na dosiahnutie cieľa práce sme použili voľne dostupné dáta z webstránky NOAA, ktorej súčasťou je databáza NGDC/WDS Global Tsunami Database¹. V nej sa zameriame na dátovú množinu obsahujúcu informácie o miestach, kde boli pozorované účinky tsunami. Ide o časové obdobie od roku 2 000 p. n. l. až po 1.6.2021, kedy sme stiahli vstupné dáta. Z geografického hľadiska ide o oblasti Atlantického, Indického a Tichého oceánu, Stredozemného a Karibského mora. Táto množina je tvorená 29 410 záznamami, ktoré sú popísané 45 atribútmi. Cieľovým atribútom v našej práci je *Tsunami Event Validity*, ktorý stanovuje platnosť tsunami pomocou šiestich kategórií: -1.0 – chybný údaj; 0.0 – udalosť, ktorá spôsobila záplavu alebo poruchu vo vnútrozemskej rieke; 1.0 – veľmi pochybné tsunami; 2.0 – pochybné tsunami; 3.0 – pravdepodobné tsunami; 4.0 – jednoznačné tsunami. Početnosť jeho hodnôt je vizualizovaná na Obr. 1.



Obr. 1 Početnosť atribútu *Tsunami Event Validity*.

Súčasťou množiny sú časové údaje iba o momente, kedy nastala príslušná zdrojová udalosť, preto ich budeme považovať za čas, kedy došlo k účinkom tsunami. Geografická poloha tsunami udalostí je charakterizovaná piatimi atribútmi: *Country*, *Area*, *Location Name*, *Latitude*, *Longitude*. Ďalšie atribúty v množine reprezentujú napríklad príčinu vzniku tsunami – *Tsunami Cause Code*, magnitúdu zemetrasenia – *Earthquake Magnitude*, vzdialenosť od zdroja – *Distance From Source (km)*, maximálnu výšku vody – *Maximum Water Height (m)*. Súčasťou sú aj atribúty obsahujúce informácie o spôsobených škodách, zväčša sú však tvorené chýbajúcimi hodnotami.

C. Príprava dát

Vzhľadom na veľkú početnosť chýbajúcich hodnôt v používanej množine sme zamerali na ich bližšiu analýzu. Pri vyjadrení ich percentuálnych podielov bolo výsledkom veľké množstvo atribútov, ktoré tvorí viac ako 80% takýchto hodnôt. Na základe zistených informácií sme extrahovali atribúty, ktoré obsahovali počet chýbajúcich hodnôt menší ako 20 000 (68%).

¹ <https://www.nci.noaa.gov/>

Podmnožina vzniknutá vyselektovaním podľa zvolených podmienok je tvorená 18 atribútmi. V rámci nej sme mapovali riadky s nulovým počtom chýbajúcich hodnôt, s menej ako dvomi, tromi a štyrmi chýbajúcimi hodnotami. V ďalších krokoch sme sa zamerali na podmnožinu tvorenú vyextrahovanými atribútmi a riadkami neobsahujúcimi chýbajúce hodnoty.

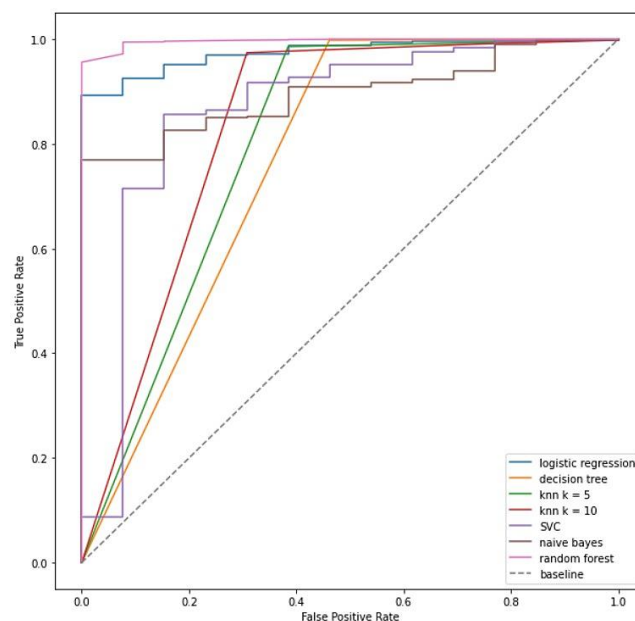
Následne sme vykonali potrebné zmeny dátových typov vrátane binarizácie a odstránili záznamy, ktoré majú v cieľovom atribúte definovanú kategóriu prislúchajúcu chybnému údaju. Zároveň sme pre uskutočnenie binárnej klasifikácie upravili rozdelenie cieľového atribútu do dvoch kategórií. Hodnoty menšie ako 4 sme zoskupili a priradili im hodnotu 0, tie budeme ďalej považovať za situácie, kedy nedošlo k tsunami udalosti. Hodnoty rovné 4 sme nahradili hodnotou 1, čo bude predstavovať validné tsunami. Početnosť jednotlivých tried v danej podmnožine je 115 zástupcov negatívnej triedy a 18 793 zástupcov pozitívnej triedy. Pre realizáciu popisného modelovania sme na numerické atribúty aplikovali normalizáciu.

D. Modelovanie a vyhodnotenie

Na riešenie binárnej klasifikácie sme použili algoritmy Logistická regresia, Rozhodovací strom, k-NN, SVM, Naive Bayes a Random Forest. Výber týchto algoritmov bol ovplyvnený poznatkami získanými v analýze prípadových štúdií z danej oblasti. Vstupné dáta boli rozdelené na tréningovú a testovaciu množinu v pomere 90:10. Ako si môžeme všimnúť v Tab. 1, väčšina modelov dosahovala vysokú senzitivitu a nízku špecificitu, čo znamená, že v mnohých prípadoch je situácia vyhodnotená ako validné tsunami, aj keď to tak nie je. Hodnoty F1 skóre len potvrdzujú neschopnosť klasifikácie negatívnej triedy.

Tabuľka 1
Výsledky modelov – 90:10 tréningová:testovacia

Model	Úspešnosť	Presnosť triedy 1	Senzitivita	Špecificita	F1 skóre		AUC
					Triedy 1	Triedy 0	
Logistická regresia	0.9931	0.9931	1.0000	0.0000	0.9966	0.0000	0.97
Rozhodovací strom	0.9947	0.9968	0.9979	0.5385	0.9973	0.5833	0.77
K-NN k = 5	0.9926	0.9931	0.9995	0.0000	0.9963	0.0000	0.80
K-NN k = 10	0.9931	0.9931	1.0000	0.0000	0.9966	0.0000	0.83
SVM	0.9905	0.9942	0.9963	0.1538	0.9952	0.1818	0.86
Naive Bayes	0.9064	0.9959	0.9095	0.4615	0.9507	0.0635	0.90
Random Forest	0.9947	0.9947	1.0000	0.2308	0.9973	0.3750	1.00

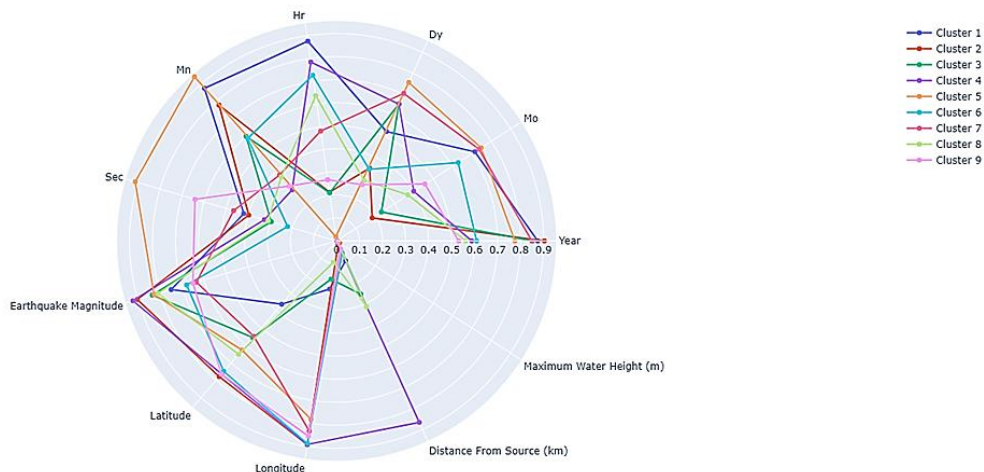


Obr. 2 ROC krivky – 90:10 tréning: testovacia

Na základe výsledkov možno povedať, že niektoré modely sú preučené, a teda klasifikujú aj príklady minoritnej triedy do majoritnej. Keďže väčšina prípadov je vyhodnotená ako situácia, kedy došlo k účinku tsunami, aj presnosť triedy 1 dosahuje vysoké hodnoty. ROC krivky modelov vizualizované na Obr. 2 len potvrdzujú, že pri tréningu modelov došlo k ich preučeniu.

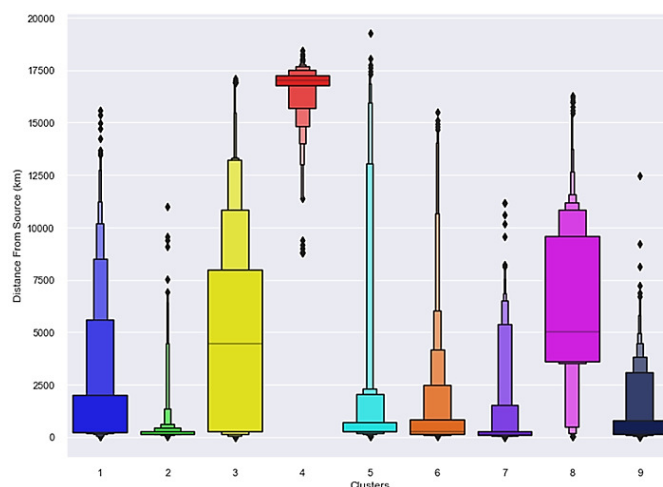
Rovnaký cyklus sme realizovali aj s použitím stratifikovanej 10-násobnej krížovej validácie a s rozdelením množiny v pomeroch 60:40, 70:30, 80:20. Vo všetkých prípadoch sa nám potvrdila najlepšia výkonnosť algoritmu Rozhodovací strom pre daný prípad. Problém preučenia modelov, s ktorým sme sa stretli v priebehu klasifikácie, sme riešili pomocou nadzorkovania dát rôznymi metódami. Na upravené dáta sme následne aplikovali rovnaké algoritmy a získané výsledky sme porovnali s pôvodnými výsledkami príslušných modelov. Vďaka tomu sme zistili, že väčšina algoritmov pri vyváženom súbore dát dosiahla vyššiu kvalitu.

V zhlukovej analýze sme aplikovali algoritmus k-Means na dáta z numerických atribútov doposiaľ používanej podmnožiny. Pre zistenie optimálneho počtu zhlukov sme využili Elbow metódu, Silhouette Coeficient a Calinski Harabasz Index. Na základe ich výsledkov sme stanovili počet zhlukov na číslo 9. Keďže jednotlivé hodnoty ich jadier prislúchajú konkrétnym atribútom, pre lepšiu prehľadnosť sme tieto hodnoty vizualizovali prostredníctvom radarového grafu na Obr. 3.



Obr. 3 Jadrá zhlukov k-Means

Najväčšie odlišnosti sme zaznamenali v atribúte *Distance From Source (km)*, ktorého distribúcia v zhlukoch je zobrazená na Obr. 4. Z vizualizácie je zjavné, že zhluk 2 obsahoval prípady hlavne s nízkymi hodnotami tohto atribútu a zhluk 4 prioritne s najvyššími hodnotami.

Obr. 4 K-Means – *Distance From Source (km)*

Na základe vykonanej analýzy možno povedať, že zhluk s najväčším počtom záznamov bol tvorený skupinou tsunami udalostí, ktoré nastali blízko ich zdroja, mali vysokú úroveň magnitúdy

zemetrásenia a vysokú maximálnu výšku vody. Oproti tomu zhuk s najmenším počtom zástupcov predstavoval nábehy, ku ktorým došlo ďaleko od zdroja, dosahovali najvyššie hodnoty magnitudy zemetrásenia a ich maximálna výška vody bola malá. K obom týmto skupinám udalostí došlo primárne na východnej strane severnej pologule. Na južnej pologuli sa predovšetkým vyskytovali prípady zo zhukov 1, 3 a 8. Všetky z nich dosahovali priemerné hodnoty magnitudy zemetrásenia. Zhuk 1 obsahoval najviac udalostí blízko zdroja z tejto trojice a zároveň najnižšie hodnoty maximálnej výšky vody. 3. zhuk mal veľké zastúpenie v širšej škále vzdialeností od zdroja a najvyššie hodnoty výšky vody z tejto skupiny. Priemerné hodnoty týchto atribútov boli súčasťou zhuku 8.

E. Nasadenie

V prvom rade je potrebné vedieť, že k účinkom tsunami dôjde. Na to je nevyhnutná ich predikcia, kde by bolo možné využiť vytvorené klasifikačné modely. Pomocou nich by sa vykonala predpoveď, či ide o údaje spôsobujúce tsunami, alebo nie. Na základe výsledku by kompetentná organizácia alebo prepojený systém mohli reagovať. Poznatky získané zo zhukovej analýzy je možné zužitkovať na predpokladanie parametrov. Napríklad, ak aj vo väčšej vzdialenosti od Japonska dôjde k zemetráseniu s priemernou hodnotou magnitudy, možno očakávať, že v tomto štáte dôjde k nábehu tsunami s malou maximálnou výškou vody. Tým pádom by bolo zjavné, že primárne sú ohrozené len pobrežné oblasti.

III. ZÁVER

Našou motiváciou bola závažnosť následkov spôsobovaných vplyvom tsunami. Celkovo sme vytvorili 131 modelov klasifikujúcich validitu účinkov tsunami. Na základe ich výsledkov sme dospeli k záveru, že najvhodnejším modelom pre daný prípad je Rozhodovací strom, keďže dokázal najlepšie klasifikovať negatívnu triedu a súčasne dosiahol dobré výsledky pre vyhodnotenie situácie, kedy nastalo tsunami. Pri zhukovej analýze pomocou algoritmu k-means sme vytvorili popisný model, na základe ktorého sme identifikovali viacero podobností, resp. odlišností, medzi údajmi. V rámci analytického procesu sme zároveň objavili niekoľko nedostatkov v dátach ako odľahlé hodnoty, nerovnomernosť tried, ktoré ovplyvňujú ich kvalitu. Za najväčší problém považujeme časové údaje nábehov tsunami, ktoré sú reprezentované časom, kedy došlo k pôsobnosti ich zdrojovej udalosti a nie nimi samými. Veľké množstvo chýbajúcich hodnôt znižuje relevantnosť viacerých atribútov a redukuje možnosti využitia dát.

POUŽITÉ ZDROJE

- [1] Y. Kusumah, B. Irawan, C. Setianingsih, *Sea wave detection system using web-based decision tree algorithm*. EEECCIS 2020 - 2020 10th Electr. Power, Electron. Commun. Control. Informatics Semin., pp. 231–236, 2020, doi: 10.1109/EECCIS49483.2020.9263444.
- [2] G. Pughazhendhi, A. Raja, P. Ramalingam, D. K. Elumalai, *Earthosys—tsunami prediction and warning system using machine learning and IoT*. vol. 28. Springer Singapore, 2019.
- [3] D. H. Dilectin, V. B. R. Mercy, *Classification and dynamic class detection of real time data for tsunami warning system*. Proc. 2012 Int. Conf. Recent Adv. Comput. Softw. Syst. RACSS 2012, pp. 124–129, 2012, doi: 10.1109/RACSS.2012.6212710
- [4] Z. Bosnjak, O. Grljevic, S. Bošnjak, *CRISP-DM as a framework for discovering knowledge in small and medium sized enterprises' data*. 2009, 509 - 514. 10.1109/SACI.2009.5136302.
- [5] R. Rivera-Lopez, J. Canul-Reich, E. Mezura-Montes, M. A. Cruz-Chávez, *Induction of decision trees as classification models through metaheuristics*. Swarm and Evolutionary Computation, Volume 69, 2022, ISSN: 2210-6502.
- [6] L. Breiman, *Random Forests*. Machine Learning, 2001 45 (1), s. 5–32.
- [7] Z. Qin, A. T. Wang, CH. Zhang, S. Zhang, *Cost-Sensitive Classification with k-Nearest Neighbors*. Cent, QCIS, Fac, Eng, Inf, Technol, Univ, Technol, Sydney, Aust., pp. 112–131, 2013,
- [8] E. Garcia-Gonzalo, Z. Fernández-Muñiz, P. J. G. Nieto, A. B. Sánchez, M. Menéndez Fernández, *Hard-Rock Stability Analysis for Span Design in Entry-Type Excavations with Learning Classifiers*. Academic Editor: Jorge de Brito, 29-06-2016.
- [9] P. Raj, A. C. Raman, H. Subramanian, *Cognitive Internet of Things: Enabling Technologies, Platforms, and Use Cases*. CRC PRESS, Boca Raton, ISBN: 978-0-429-32822-0.
- [10] D. G. Kleinbaum, M. Klein, *Logistic Regression A Self-Learning Text Third Edition*. SPRINGER, New York Dordrecht Heidelberg London, ISBN: 978-1-4419-1742-3.
- [11] R. Gustriansyah, N. Suhandi, F. Antony, *Clustering optimization in RFM analysis based on k-means*. Faculty of Computer Science, Universitas Indo Global Mandiri, Indonesia, 2020, pp. 470–477, doi:10.11591/ijeecs.v18.i1.pp470-477
- [12] H. R. Arabnia, K. Ferens, D. de la Fuente, E. B. Kozerenko, J. A. O. Varela, F. G. Tinetti, *Advances in Artificial Intelligence and Applied Cognitive Computing Proceedings from ICAI'20 and ACC'20*. SPRINGER NATURE SWITZERLAND AG, 2021, ISBN: 978-3-030-70296-0

Robot Control in the Environment of Virtual Reality

Ján VAŠČÁK

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

jan.vascak@tuke.sk

Abstract – The aim of this paper is to present the problem of remote control of a mobile robot using virtual reality. In order to address this issue, the thesis deals with the creation of a virtual environment, dynamic insertion of objects into such an environment, its modeling as well as the design of an algorithm for robot navigation. Therefore, this thesis theoretically describes concepts such as virtual reality, augmented reality as well as the ways in which this task can be made. Finally, the thesis deals with the actual implementation of the design and the simulation of the developed solution. Subsequently, the performed experiments and their evaluation are described.

Keywords – driving, navigation, robot, virtual reality

I. INTRODUCTION

Nowadays, the robot is a common technology in almost every household as well as in companies, whether they are smaller or larger enterprises. One of the biggest dilemmas in the world of robotics is how to properly navigate a robot in space.

The aim of this work was to model a virtual model of a robot, insert it into a virtual environment and then build software for its navigation in this virtual environment as well as inserting objects to be avoided.

II. VIRTUAL REALITY

The term virtual reality (VR) has been defined in many different ways (e.g. [1]), which has caused that currently VR can mean different things in different contexts. VR can range from a simple environment presented on a desktop computer to a fully immersive multisensory environment containing hardware, such as special clothing or special goggles for the head [2], [3].

In all its manifestations, VR is seen as a way to simulate or replicate an environment and consequently offer the user the feeling of being in that environment, taking control and personally interacting with that environment using their own body.

In addition to simulating a three-dimensional environment, all forms of VR share common input as well as computer control. It is widely known that the true essence of VR lies in its computer-generated 3D worlds. The VR interface immerses participants in a 3D synthesized environment that is created by one or more computers, thus allowing them to act in real time in this environment using one or more control devices and involving one or more of their physical senses. Virtual reality results in stimulation of the participant's senses, giving the vivid impression of being immersed in a synthetic environment with which the participant is interacting.

Based on several years of experience, it can be argued that VR primarily finds its use as an instructional technology, e.g. medical or dental education. Engineering education into the field supporting the effectiveness of VR as an instructional tool is also finding its way Ausburn.

Recently, there has been a tendency for research to focus on so-called immersive virtual reality more than on computer virtual reality. While virtual reality has advantages as an instructional technology, researchers have also pointed out its limitations. One important limitation is the high level of skill and cost required to develop and implement VR, especially when it is an immersive virtual reality.

The development of immersive VR requires a very high level of expertise in programming and graphics as well as very expensive hardware and software to use it effectively. Although new VR technology for desktop computers has dramatically reduced the skill requirements and costs for the necessary hardware, significant financial as well as time investment is still required.

Another set of limitations of VR arise from the nature of the equipment that is a necessity. A long term problem with helping VR is the health and safety concerns from a long term use perspective, e.g. headaches, nausea, balance disturbances and other physical problems caused by the use of these systems. Although these problems have largely disappeared from current VR research, as equipment has improved in the meantime and improvements appear to have been made in new desktop computer systems, little is still known about the long-term physical or psychological effects of VR use.

III. AUGMENTED REALITY

Augmented reality can be defined as a direct or indirect view of a real-world physical environment in real time that has been enhanced or augmented with information generated by a virtual computer [3].

Augmented reality is interactive and registered in a 3D environment and combines both real and virtual objects. The so-called Milgram real-virtual continuum is defined as a continuum whose span between the real environment and the virtual environment includes augmented reality (AR) and augmented virtuality (AV), where AR tends closer to the real world and AV tends closer to the virtual environment.

The main goal of augmented reality is to simplify the user's life by bringing virtual information not only to the user's immediate surroundings, where they are, but also to any indirect view of the environment in the real world, such as a live video stream. Augmented reality enhances the user's perception of and interaction with the real world. While VR technology completely immerses the user in a synthetic world without seeing the real world, AR technology enhances the sense of reality by inserting virtual objects into the environment and hinting at the real world in real time.

AR cannot be limited to a particular type of imaging technology, such as HMDs, or to the visual senses alone. Augmented reality can potentially be applied to all other senses such as smell, touch or hearing. Augmented reality can also be used as an augmentation or replacement, if the user has lost a sense, by sensory substitution, such as augmenting vision for blind or low vision users with audio cues, or augmenting hearing for deaf users with visual cues.

An approach that modifies AR for applications that require removal of real objects from the environment is also possible. In this case, AR is replaced by a name such as *mediated reality* or *reduced reality*, where the removed objects are replaced by virtual objects [4]. Removing objects from the real world is the real-world equivalent of covering the object with virtual information that matches the background to give the user the impression that the object is not there. Virtual objects added to the real environment show the user information that the user cannot directly detect using his/her senses.

The information conveyed by virtual objects can help the user to perform everyday tasks, such as guiding pilots in an airplane where digital information is displayed via a headset. The information can also be used for tracking, such as in *Wikitude* or other augmented reality mobile applications. There are many other classes of applications using AR, such as medical visualization, entertainment, advertising, maintenance and repair, annotation, or robot navigation.

IV. SOFTWARE AND SIMULATION IMPLEMENTATION OF A ROBOT CONTROL SYSTEM

At the beginning it is necessary to define in which game engine, i.e. programming environment and programming language, the development of robot control in the virtual environment will be realized. The game engine, as mentioned in the design, will be *Unity3D*. As a programming language for programming algorithms, the programming language *C#* will be used and the programming environment for this language will be Visual Studio 2019 Community, which is very compatible with this programming language as well as with the *Unity3D* game kernel. the whole system can be divided into front end and back end.

Front end applications are used to visualize the activity of work, which is controlling the robot in a virtual environment. The front end uses elements that are contained in the *Unity3D* kernel itself, such as a camera, lighting, and various 3D objects such as a cube or a plane. The front end visualizes the scene, which contains the lab model itself, the robot (in our case Turtlebot), also its movement to the target destination, the cursor pointer, the destination marked by the user, or the embedded objects.

The front end contains primarily the *MainScene* component, which provides a repository for all other components to be used in a given application view. As such, this type of component has no properties to work with. The *MainScene* is shown in Fig. 1 and contains these other components such as *LabLight*, *TurtleBot*, *Manager*, *WorldCamera*, *Lab*, *Terrain* and *WorldLight*.

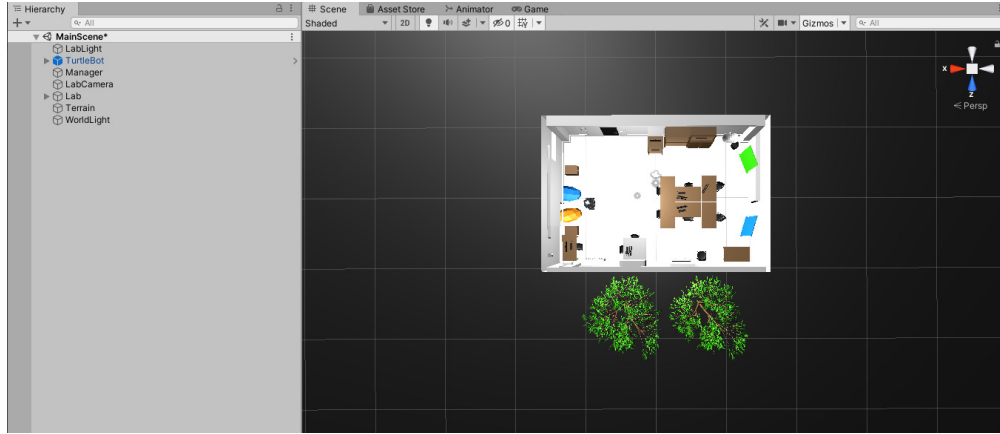


Fig. 1 View of the *MainScene* containing components.

The back end contains several basic components whose main role is to process data, including the generation of control signals for the robot. In addition, it also performs some support tasks for the front end itself. Fig. 2 shows a basic relationship diagram between these components, and we will briefly describe some of them next.

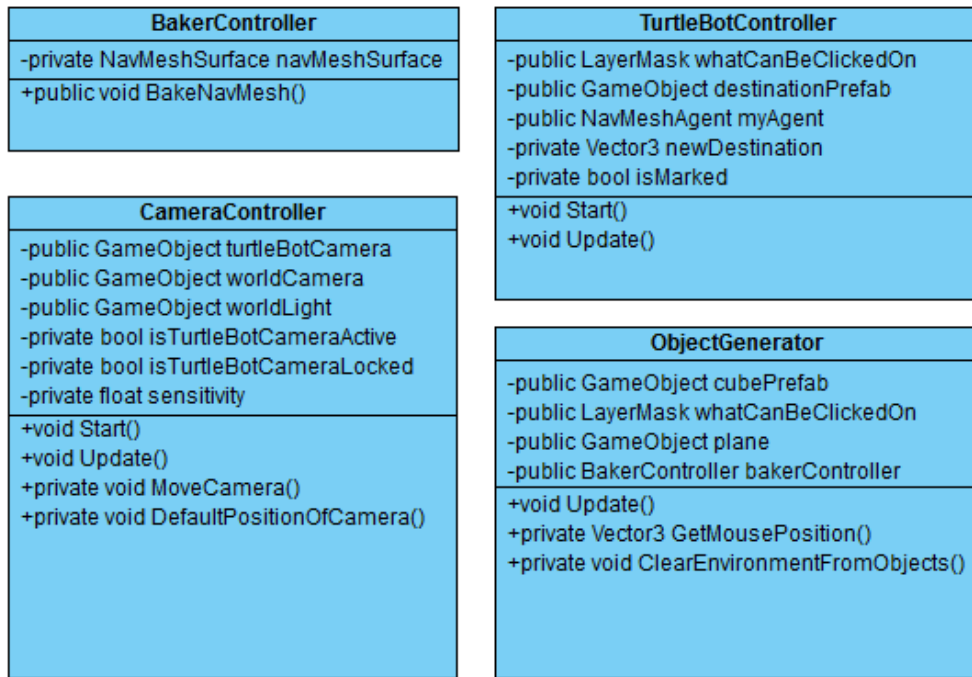


Fig. 2 Diagram of robot control components in a virtual environment.

The *BakerController* component is used to "bake" the surface, i.e. it detects which places on the surface are passable by the robot and which are not. This function, as can be seen in Fig. 2, contains one private global variable *navMeshSurface*, which represents the front end component *Plane*, and a public method *BakeNavMesh()*, which includes the code for this operation. The *CameraController* is responsible for providing functions for all cameras that are in the scene. It provides most of the functionality for the camera that is deployed on the mobile robot. The *ObjectGenerator* allows to insert objects into the virtual environment the robot is in. Finally, the *TurtleBotController* component represents the actual navigation and motion control of the robot *TurtleBot* [5], which uses the functionality offered by the *Unity.AI* library.

V. EXPERIMENTS AND THEIR EVALUATION

In order to verify the functionality of the proposed system, two types of experiments were conducted, namely shadow experiments and embedded object experiments.

Table 1
 Table of the maximum number of FPS created in the shadows experiment.

Experiment	Shadows	Objects	FPS
1.	Off	No	29
2.	On	No	18
3.	Off	Yes (10 pcs)	29
4.	On	Yes (10 pcs)	18
5.	Off	Yes (20 pcs)	29
6.	On	Yes (20 pcs)	18

The shadow experiment focuses on the effect of shadows on the image fluency of robot control in a virtual environment. First, robot control experiments are performed with shadows on off, then robot control experiments are performed with shadows on.

Table 1 shows the data obtained by performing individual experiments with different parameters. The first two experiments were performed first without embedded objects, where only the shadows were turned off and on. It can be seen that turning on the shadows has a large effect on the smoothness of the image. The subsequent next four experiments were also performed with inserted objects. First, 10 pieces of objects were inserted, variously distributed in the virtual environment, and the achievability of the maximum number of frames per second (FPS) created with shadows on and off was then determined.

This experiment showed that the image is smoother when the shadows are turned off than when they are turned on. The number of embedded objects is not significant in increasing or decreasing the FPS value.

The embedded object experiment looks at the time taken for the robot to move from the starting point to the prescribed destinations in the virtual environment, which are shown by the crosses in Fig. 3. At each destination, the robot will be in a situation where it has a clear path with no obstacles and in a situation where objects will be placed in its path.

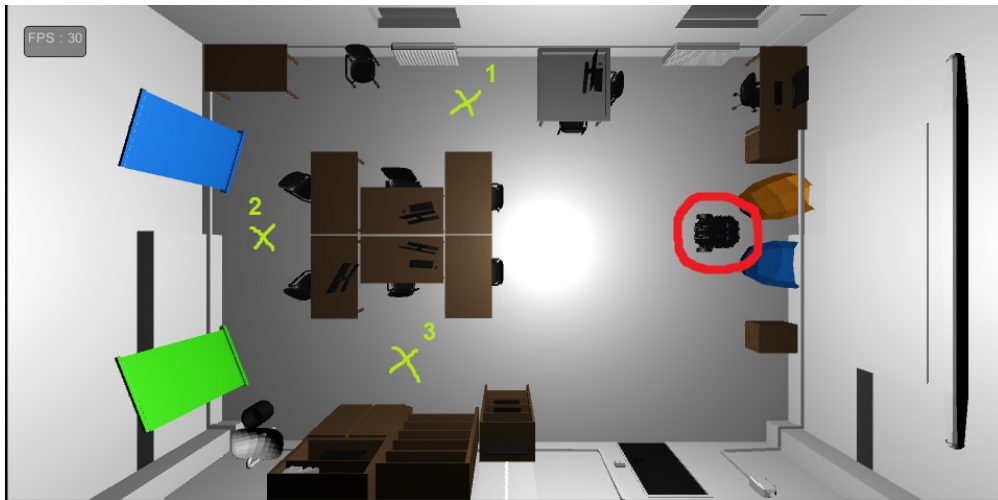


Fig. 3 Destinations used in the virtual environment denoted by x.

Fig. 4 shows the experiments performed. In the upper left corner are the numbers of the experiments. As can be seen in Table 2, the robot achieved the shortest time when moving without obstacles to position 1. In the situation where virtual objects were placed in front of the robot, the robot arrived at the worst time when moving to position 1. The lowest difference in time for both the obstacle-free and with obstacles transfer was at position 2, where the difference was only less than a second.

VI. CONCLUSIONS

The aim of this work was to model a virtual robot, a virtual space, to develop the control of this robot in a virtual environment, as well as to develop a function to insert objects into the virtual environment.

The experimental results confirm the success of this proposal. This work has the potential for future extension as mobile robots are increasingly used in industrial manufacturing and it

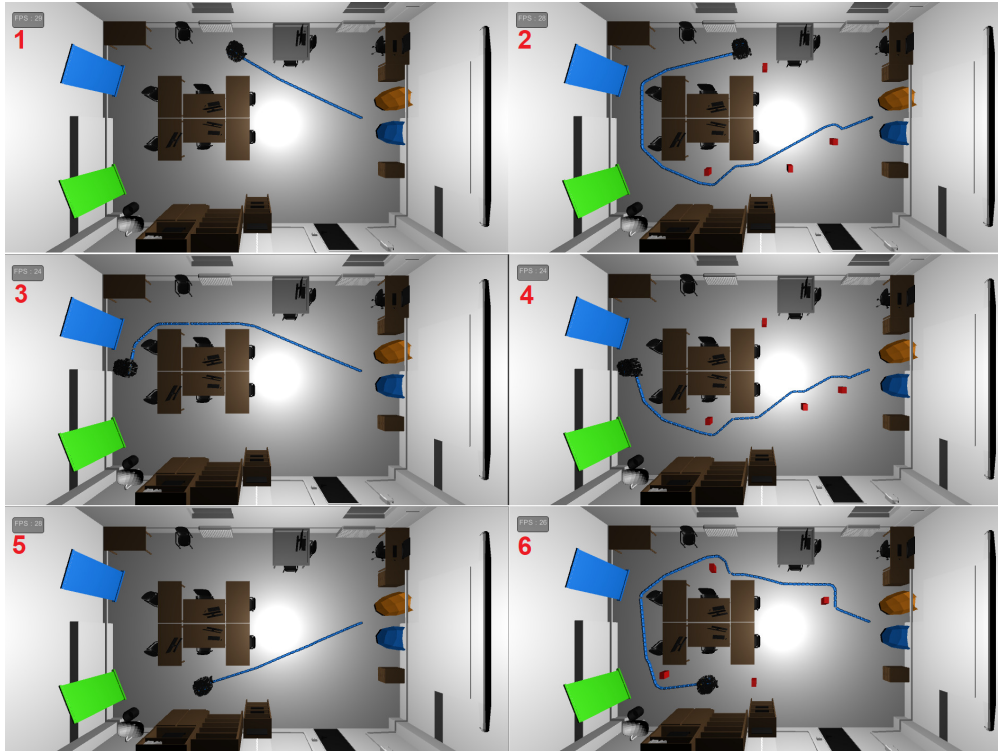


Fig. 4 Experiments shown in a virtual environment.

Table 2
 Table of data obtained in the object embedding experiment.

Experiment	Destination	Obstacles	Time [sec.]
1.	1	0	5.19
2.	1	4	16.31
3.	2	0	9.81
4.	2	4	10.76
5.	3	0	6.09
6.	3	4	15.17

is necessary for the robot to be able to navigate even if it encounters an unexpected obstacle while navigating. In this work, the *RosSharp* library is implemented, which in the future offers the possibility to link a virtual robot model to a real robot by connecting to its IP address and solving tasks such as environment mapping [6].

ACKNOWLEDGMENT

This publication is the result of the APVV grant ENISaC – Edge-Enabled Intelligent Sensing and Computing (APVV-20-0247).

REFERENCES

- [1] A. Alrazzah, F. Lamya, and L. Fattouh, “Environments and system types of virtual reality technology in STEM: A survey,” *International Journal of Advanced Computer Science and Applications*, vol. 8, pp. 77–89, 06 2017.
- [2] L. J. Ausburn and F. B. Ausburn, “Desktop Virtual Reality: A Powerful New Technology for Teaching and Research in Industrial Teacher Education,” *Journal of Industrial Teacher Education*, vol. 41, no. 4, pp. 1–16, 2004.
- [3] J. Carmigniani, B. Furht, M. Anisetti, P. Ceravolo, E. Damiani, and M. Ivkovic, “Augmented reality technologies, systems and applications,” *Multimedia Tools Applications*, vol. 51, no. 1, pp. 341–377, Jan. 2011.
- [4] M. R. Marner, R. T. Smith, J. A. Walsh, and B. H. Thomas, “Spatial user interfaces for large-scale projector-based augmented reality,” *IEEE Computer Graphics and Applications*, vol. 34, no. 6, pp. 74–82, 2014.
- [5] A. d. H. B. M. Tavares, S. P. Madruga, A. Brito, and T. P. Nascimento, “Non-linear model predictive formation control applied to non-holonomic autonomous mobile robots,” in *2017 Latin American Robotics Symposium (LARS) and 2017 Brazilian Symposium on Robotics (SBR)*, 2017, pp. 1–6.
- [6] H. Aagela, M. Al-Nesf, and V. Holmes, “An asus_xtion_probased indoor mapping using a raspberry pi with turtlebot robot,” in *2017 23rd International Conference on Automation and Computing (ICAC)*, 2017, pp. 1–5.

Univerzálny emulátor energetických systémov

¹Marek FEDOR, ²Daniela PERDUKOVÁ

^{1,2} Katedra elektrotechniky a mechatroniky, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

¹marek.fedor.4@student.tuke.sk, ²daniela.perdukova@tuke.sk

Abstract — Výskum, optimalizácia a praktická realizácia procesov optimalizácie energetických podsystémov elektrární, teplární a priemyselných podnikov je pomerne zložitá úloha, ktorá je dnes nepredstaviteľná bez silných špecializovaných nástrojov počítačovej podpory. V článku je uvedený okrem návrhu koncepcie emulátora energetických systémov aj návrh modelu hydraulického turbíny ako najdôležitejšieho podsystému malej vodnej elektrárne, ktorého výpočet účinnosti je analyticky náročný a závislý od parametrov, ktoré sa často získavajú len teoretickým odhadom. V tomto článku bol na vytvorenie jeho modelu použitý fuzzy systém na základe nameraných prevádzkových údajov prietoku turbíny a výšky jej vodného stĺpca bez potreby presnej znalosti jej parametrov. Samotný emulátor energetických systémov je nástroj určený pre vývojové a aplikačné práce s energetickými systémami, ktorý je postavený na báze dnes štandardne používaných decentralizovaných riadiacich systémov (DCS). V článku sú uvedené základné princípy modulárnej stavby emulátora a príklad jeho realizácie na báze DCS systému SIMATIC-S7 od firmy Siemens.

Keywords — modelovanie energetických procesov, decentralizovaný riadiaci systém DCS, emulátor na báze DCS

I. ÚVOD

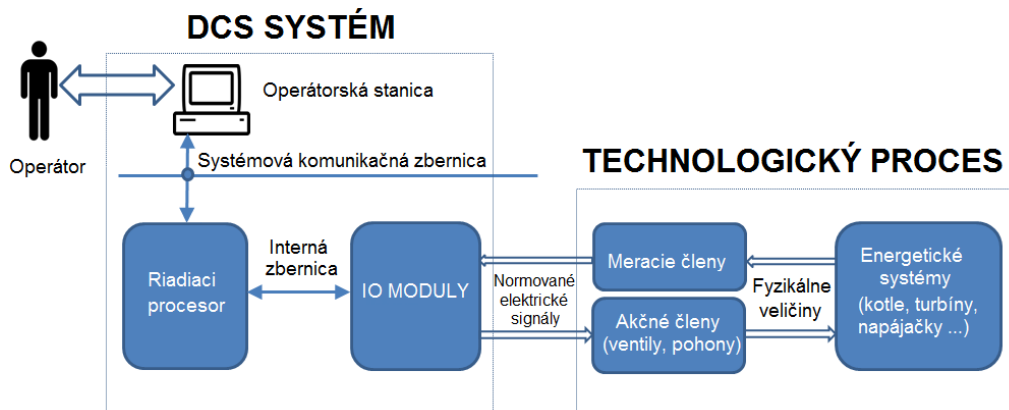
V súčasnosti sa riadenie výroby a distribúcie elektrickej energie energetických podnikov (elektrární, teplární a energetiky vo veľkých priemyselných podnikoch) realizuje v prevažnej miere na báze konkrétnych riešení a vylepšovaním existujúcich algoritmov na základe priebežne získavaných praktických skúseností. Často len veľké nadnárodné firmy majú svoje výskumné tímy a metódy, ktoré sú však mnohým podnikom a firmám nedostupné z dôvodu finančnej náročnosti. Cieľom tohto príspevku je navrhnúť koncepciu univerzálneho energetického emulátora, ktorý by čo najviac odpovedal konkrétnemu reálnemu energetickému systému. Ide o zariadenie, ktoré by umožňovalo modelovo čo najvernejšie v súlade so skutočným systémom skúmať a navrhovať konkrétne podsystémy elektrární, teplární a ďalších výrobných procesov. Dosiahnuté znalosti by bolo možné ďalej priamo využiť pri praktickom nasadení konkrétneho riadiaceho systému, a to hlavne pri riadení takýchto systémov [1-2], ich vizualizácii [3] a tiež pri zaškolení obsluhujúceho personálu (operátorov turbín, kotlov, malých vodných elektrární a pod.).

II. KONCEPCIA EMULÁTORA ENERGETICKÝCH SYSTÉMOV

Koncepcia emulátora vychádza z požiadaviek na jeho využívanie v technickej praxi podnikov a firiem, ktoré sa zaoberajú nasadzovaním riadiacich systémov pre energetické a teplárenské procesy [4]. Schéma riadeného energetického systému, ktorý je potrebné emulovať, je ukázaná na Obr.1.

Energetický systém má dve hlavné časti, a to technologickú a riadiacu. Reálne emulovať riadiacu časť energetického systému iným (jednoduchším, lacnejším a pod.) počítačovým systémom ako je decentralizovaný systém (DCS) je prakticky nemožné. Z uvedeného dôvodu je optimálne a najvýhodnejšie postaviť technické prostriedky navrhovaného emulátora na báze štandardizovaného DCS systému. Časť emulátora bude simulovať samotný DCS podsystém a druhá časť energetický proces s jeho podsystémami. Toto riešenie ďalej umožňuje emulátorom navrhnuté regulačné a vizualizačné bloky priamo aplikovať a nasadiť pre riadenie konkrétne skúmaného energetického zariadenia. Tento koncept takisto umožňuje školiť operátorov zariadení prakticky na reálnom HMI (Human Machine Interface), čo podstatne zvyšuje kvalitu ich získaných vedomostí a skúseností. Ďalšou výhodou tohto riešenia je

možnosť testovania konkrétneho zapojenia jednotlivých meracích a akčných členov navrhovaného systému priamo na emulátore ešte pred ich projektovým návrhom. čo znižuje množstvo potrebných zmien projektu pri implementácii a tiež značne urýchľuje dobu realizácie konkrétneho energetického systému v praxi.



Obr.1 Názorná schéma riadeného energetického systému

Vzhľadom na požiadavku univerzálnosti navrhovaného emulátora je nutné jeho programové vybavenie (software) stavať hierarchicky a prísne modulárne. Jedná sa o tri základné typy modulov:

Riadiace moduly. Najpoužívanější typy takýchto modulov sú dnes štandardnou súčasťou každého DCS systému. Jedná sa o klasické PID regulátory, binárne regulátory pohonov a ventilov rôzneho typu, sekvenčné regulátory a pod. V prípade potreby iných, menej štandardných regulátorov (fuzzy regulátory, neurónové siete, regulátory s referenčným modelom, adaptívne regulátory a pod.) je potrebné vytvoriť v DCS systéme nový knižničný modul na základe štandardov použitého systému.

Intefejsové moduly. Tento typ modulov je dnes takisto štandardnou súčasťou každého DCS systému a je určený pre prepojenie tohto systému s príslušnou technológiou. Pretože u navrhovaného emulátora budú tieto moduly slúžiť prevažne na interné prepojenie s technologickými modulmi, je potrebné ich upraviť, a to hlavne z hľadiska jednoduchšej možnosti presmerovania ich vstupných a výstupných signálov.

Technologické moduly. Tieto moduly sú jadrom emulátora a je potrebné ich vyvinúť, verifikovať a realizovať pre každý jednotlivý podsystem energetického zariadenia najvšeobecnejšou formou.

III. REALIZÁCIA EMULÁTORA ENERGETICKÝCH SYSTÉMOV NA BÁZE DCS SYSTÉMU FIRMY SIEMENS

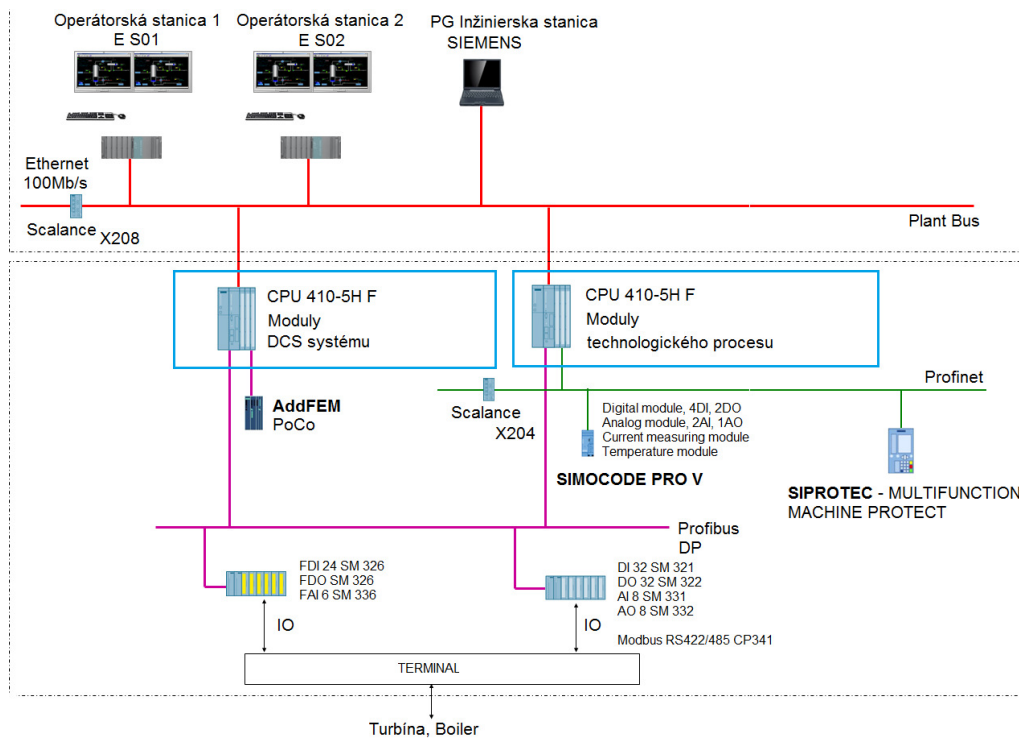
Na základe koncepcie emulátora energetických systémov z Obr. 1 bolo navrhnuté a zrealizované HW vybavenie emulátora na báze systému SIMATIC-S7 firmy Siemens podľa Obr.2.

Celý emulátor je postavený na dvoch procesoroch, z ktorých jeden emuluje samotné riadenie a druhý energetické technologické zariadenie. Súčasťou emulátora sú dve operátorské stanice (môžu slúžiť zároveň aj ako inžinierske) pre vizualizáciu a jedna prenosná programovacia stanica. Procesory sú navzájom prepojené cez zbernicu Profibus DP. Pri emulácii a školení operátorov budú všetky signály medzi riadiacou a technologickou časťou emulovaného zariadenia (teda procesorom 1 a procesorom 2) prenášané cez túto zbernicu. Po pripojení na reálne zariadenie budú signály z modulov prepnuté na reálne IO karty systému.

Programové vybavenie (softvér) emulátora je postavený na programových moduloch, zodpovedajúcich jednotlivým podsystemom energetického zariadenia.

Technologické moduly sú realizované podľa konkrétneho technologického podsystemu. Každý takýto modul má tieto časti:

Jadro modulu. Jedná sa o programový blok, v ktorom je naprogramovaná funkčná závislosť medzi relevantnými vstupmi a výstupmi technologického podsystemu. Táto závislosť môže byť modelovaná explicitne analyticky (pozri obr.3) alebo implicitne (napr. pomocou fuzzy pravidiel, popisujúcich relácie medzi vstupmi a výstupmi).



Obr.2 Bloková schéma emulátora na báze DCS firmy Siemens

Blok parametrov modulu. Vzhľadom na požadovanú univerzálnosť emulátora je zrejmé, že rozdielnosť dvoch energetických blokov rovnakého typu bude daná množinou ich vnútorných parametrov, preto každý technologický modul musí mať podsystem pre evidenciu svojich parametrov a formu ich nastavovania. Toto môže byť problémom hlavne u nelineárnych systémov, pretože ich nelinearity nemusia byť explicitne matematicky popísané. Pre daný konkrétny technologický prvok (napr. turbínu) je možné ich získať meraním a zobrazit' pomocou tabuľky alebo vhodnej FIS (Fuzzy Inference System) štruktúry.

I/O interface modulu. Tento interface slúži na prepojenie technologického modulu s inými modulmi, prípadne s okolím systému DCS. Okrem spracovania IO signálov musí obsahovať aj možnosť presmerovania signálov zo svojej IO tabuľky na definované HW IO karty.

Vizualizačný interface modulu. Na rozdiel od riadiacich modulov, ktoré prijímajú signály (povely, prepínanie módu a pod.) od operátora (resp. nadradených modulov) a aj vizualizujú svoje stavy, technologické moduly je potrebné iba vizualizovať (veľmi zriedka nastaviť nejaký prepnutý stav). Na druhej strane má energetický podsystem bežne na vizualizáciu podstatne viac signálov ako štandardný regulačný blok. Preto tento interface obsahuje hlavne vizualizačné obrázky pre daný technologický prvok.

IV. MODEL TURBÍNY MALEJ VODNEJ ELEKTRÁRNE

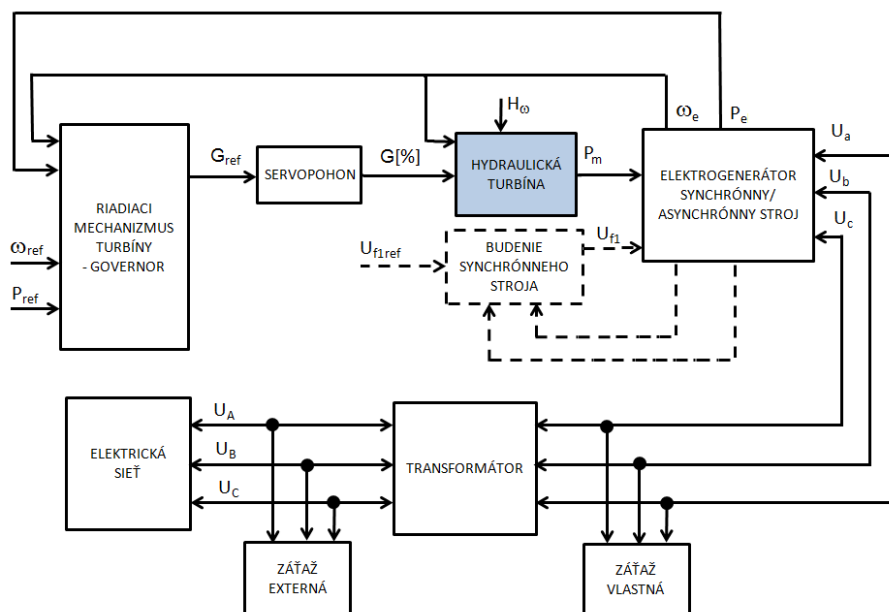
Ako príklad energetického zariadenia je možné uviesť malú vodnú elektrárňu.

Vo všeobecnosti sa simulačný model malej vodnej elektrárne (MVE) skladá z piatich hlavných podsystemov, a to:

- governora – predstavujúceho systém riadenia turbíny;
- servopohonu, ktorý slúži ako pohon na reguláciu prietoku vody cez hydraulickú turbínu pomocou ovládania ventilov alebo rozvádzačích lopatiek podľa výstupného signálu Governora;
- hydraulickej turbíny – predstavujúceho proces premeny energie;
- bloku elektrického generátora a elektrickej siete.

Na Obr.3 sú znázornené všetky vyššie uvedené hlavné podsystemy tvoriace základný simulačný model MVE.

Hydraulická turbína je kvôli jej dôležitej funkcii v celom procese výroby elektrickej energie považovaná za srdce vodnej elektrárne. Turbína premieňa kinetickú alebo potenciálnu energiu vody na mechanickú rotáciu, ktorá pohybuje rotorom generátora cez spoločný hriadeľ. Na dosiahnutie realistických a relevantných výsledkov je preto potrebné použiť jej presný model.



Obr.3 Bloková schéma malej vodnej elektrárne MVE

V. MODEL PODSYSTÉMU HYDRAULICKEJ TURBÍNY

Obrázok 4 zobrazuje blokovú schému podsystemu hydraulickej turbíny, ktorý bol vytvorený na základe teoretického a matematického popisu turbíny uvedeného v literatúre [5].

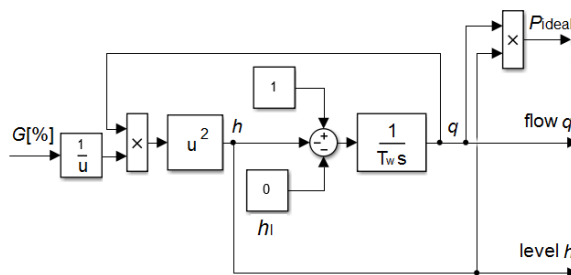


Fig.4 Bloková schéma modelu vodnej turbíny

Modelovanie prívodového potrubia, ako súčasť znázorneného modelu, bolo robené za predpokladu nestlačiteľnosti vody a za predpokladu, že prívodové potrubie je dostatočne pevné s dĺžkou L a plochou prierezu A . Úbytok výšky vodného stĺpca následkom vplyvu prívodového potrubia je zanedbaný.

Rovnica 1 predstavuje základný vzorec použitý pre návrh simulačného modelu hydraulickej turbíny.

$$\frac{d\bar{q}}{dt} = (\bar{h}_0 - \bar{h} - \bar{h}_l)g \frac{A}{L} \quad (1)$$

Tabuľka 1. Zoznam parametrov použitých v rovnici (1)

Symbol	Popis	Jednotka
q	Prietok turbíny	$[\text{m}^3/\text{s}]$
A	Plocha prierezu prívodného potrubia	$[\text{m}^2]$
L	Dĺžka prívodného potrubia	$[\text{m}]$
g	Gravitačné zrýchlenie	$[\text{m}/\text{s}^2]$
h_0	Výška vodného stĺpca	$[\text{m}]$
h	Výška vodného stĺpca na vstupe do turbíny	$[\text{m}]$
h_l	Strata výšky vodného stĺpca spôsobená trením v prívodnom potrubí	$[\text{m}]$

A. Model hydraulické turbíny zohľadňujúci straty v potrubí

V reálnej prevádzke hydraulická turbína nepracuje so 100% účinnosťou a z toho vychádza požiadavka na zohľadnenie tejto skutočnosti aj pri jej modelovaní. Spôsobov, ako zakomponovať vplyv účinnosti turbíny do simulačného modelu je viacero [5-6].

Jedným z najjednoduchších spôsobov je použitie takzvaného prietoku pri nulovom zaťažení. Tento prietok označovaný ako q_{nl} je konštantou, charakterizujúcou hodnotu stálych výkonových strát turbíny, pričom vo vzorci je táto konštanta odčítaná od aktuálneho prietoku turbínou, a až táto výsledná hodnota prietoku je prenasobená výškou vodného stĺpca h podľa nasledujúcej rovnice

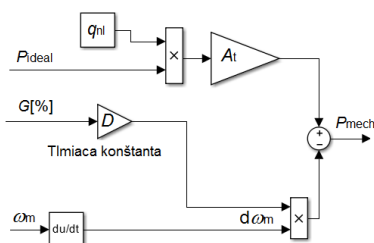
$$P_m = A_t h (q - q_{nl}) - DG \Delta \omega \quad (2)$$

Rovnako sa počas prevádzky elektrárne uplatňuje aj tlmiaci efekt turbíny, ktorý je funkciou rozdielu rýchlostí, polohy servopohonu a polohy ventilu prívodu vody. Veličina $\Delta \omega$ reprezentuje zmenu mechanickej rýchlosti hriadeľa turbíny, D je veľkosť priemeru prívodového potrubia (tiež ako tlmiaci koeficient turbíny) a G je poloha servopohonu v percentách.

Parameter A_t je považovaný za zosilnenie hydraulické turbíny. V rôznych vedeckých publikáciách sa jeho vyčíslenie líši, napr. podľa literatúry [7] ho môžeme vypočítať nasledovne

$$A_t = \frac{\text{Výkon Turbíny [MW]}}{(\text{Výkon Generátora [MVA]} \cdot h_r (q_r - q_{nl}))} \quad (3)$$

pričom veličiny s dolným indexom r môžeme považovať za nominálne (rated). To znamená, že q_r je prietok vody turbínou pri jej nominálnom zaťažení a h_r je výška vodného stĺpca potrebná na dosiahnutie spomínaného nominálneho prietoku q_r . Zvyčajne sú tieto hodnoty súčasťou technickej dokumentácie jednotlivých systémov vodných elektrární. Model výpočtu strát hydraulické turbíny je zobrazený na Obr.5.



Obr.5 Model výpočtu strát hydraulické turbíny

B. Model zohľadňujúci účinnosť turbíny prostredníctvom výpočtu jej účinnosti

Tento variant tvorby simulačnej štruktúry hydraulické turbíny vychádza zo základného vzťahu popisujúceho výpočet jej výstupného mechanickeho výkonu P_m

$$P_m = \eta \rho q g h \quad (4)$$

kde η predstavuje hodnotu účinnosti hydraulické turbíny; ρ je hustota vody (zvyčajne sa používa hodnota 1000 kg/m^3); q je aktuálna hodnota toku vody turbínou; g je gravitačné zrýchlenie ($9,81 \text{ m/s}^2$); a h je aktuálna výška vodného stĺpca vodnej elektrárne. V modeli sú použité normované veličiny. V tomto prípade je na vyjadrenie účinnosti hydraulické turbíny použitý výpočet, ktorý je uvedený v literatúre [8-9], pre ktorý je potrebné poznať parametre turbíny ako polomer lopatiek turbíny - R_{blades} [m] a veľkosť plochy, ktorú dané lopatky „obtiarajú“ - A_{blades} [m²]. Taktiež sú pre vyčíslenie nevyhnutné aj aktuálne hodnoty toku vody turbínou q , ako aj jej mechanickej uhlovej rýchlosti ω_m

$$\eta(\lambda, q) = \left[\frac{1}{2} \left(\frac{90}{\lambda_i} + q + 0.78 \right) \exp \left(\frac{-50}{\lambda_i} \right) \right] (3,33q) \quad (5)$$

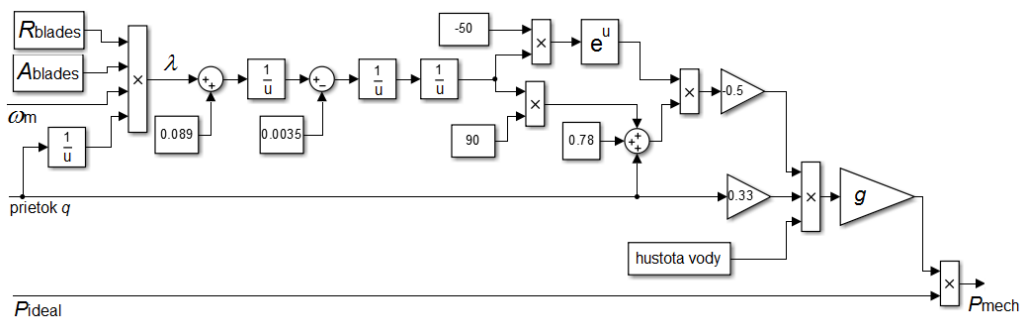
kde pre λ_i platí

$$\lambda_i = \left[\frac{1}{(\lambda + 0,089)} - 0,0035 \right]^{-1} \quad (6)$$

a λ je definovaná ako

$$\lambda = \frac{R_{blades} A_{blades} \omega_m}{q} \quad (7)$$

Vstupom do výpočtu účinnosti hydraulické turbíny by bol v tomto prípade aktuálny prietok vody turbínou q a mechanickej uhlovej rýchlosti turbíny ω_m . Matematickému popisu zodpovedá bloková schéma na Obr.6.

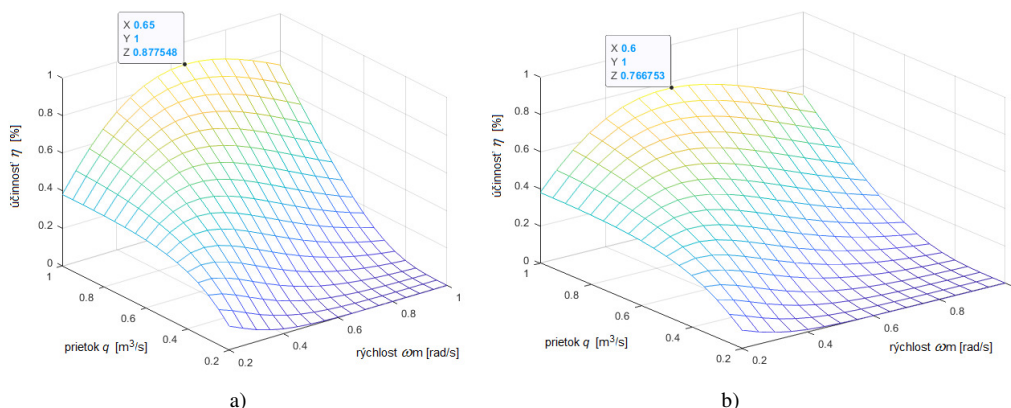


Obr.6 Model výpočtu účinnosti hydraulické turbíny

VI. FUZZY MODEL ÚČINNOSTI HYDRAULICKEJ TURBÍNY

Všetky vyššie uvedené spôsoby výpočtu účinnosti turbíny sú založené na pomerne komplikovaných a nie vždy presných výpočtoch, a to najmä z dôvodu potreby detailnej znalosti jej parametrov. Napr. podľa rovnice (7) závisí účinnosť od parametrov lopatiek konkrétnej turbíny, ktoré sa počas prevádzky môžu výrazne meniť. Táto závislosť je ďalej silne nelineárna (rovnice 5-7) a závislá od viacerých približne určených parametrov. Napr. pre polomer lopatiek $R_{blades} = 1.9$ m a plochu ich obtierania $A_{blades} = 9$ m² je ukázaná účinnosť pre rôzne prietoky a mechanické uhlové rýchlosti turbíny na Obr.7a.

Pri zmene plochy obtekania lopatiek (napr. opotrebovaním) na $A_{blades} = 8$ m² sa tento „obraz účinnosti“ turbíny zmení podľa Obr.7b. Maximálna účinnosť turbíny v tomto prípade je približne 76,6%, zatiaľ čo v predchádzajúcom prípade bola 87,7%.



Obr.7 Účinnosť turbíny a) $R_{blades}=1.9$ m a $A_{blades}=9$ m², $\eta=87,7\%$
 b) $R_{blades}=1.9$ m a $A_{blades}=8$ m², $\eta=76,6\%$

Na Obr. 7 je zrejмый rozdiel v maximálnej možnej hodnote účinnosti turbíny, spôsobený len opotrebovaním lopatiek turbíny, čo má za následok zmenu vnútornej plochy lopatiek (parameter A_{blades}). Tento parameter má podľa rovnice (7) významný vplyv na výpočet účinnosti hydraulické turbíny, a preto je problematické zohľadňovať jeho vplyv pri tvorbe jej simulačného modelu.

Ďalšie kroky nášho výskumu v oblasti modelovania podsystemu hydraulické turbíny sú v zvažovaní ďalších možností z hľadiska univerzálneho, no stále presného a rýchleho výpočtu jej účinnosti.

Obraz účinnosti turbíny je možné vytvoriť aj na základe merania prietoku vody turbínou q , mechanickej uhlovej rýchlosti hriadeľa ω_m a účinnosti turbíny η v ustálených prevádzkových stavoch, tak ako je to uvedené v Tab.2.

Údaje z Tab. 2 možno použiť na vytvorenie fuzzy inferenčnej štruktúry (FIS) modelovanej účinnosti popisujúcej nameraný vzťah medzi $[\omega_m, q] \rightarrow \eta$.

Na základe týchto údajov bol navrhnutý fuzzy model účinnosti pomocou štandardných nástrojov zhlukovej analýzy, ktorý vedie k zlepšeniu kvality simulačného modelovania a k skráteniu času potrebného na tvorbu fuzzy modelu [10]. Základnou charakteristikou zhlukovej analýzy je zníženie celkového počtu fuzzy pravidiel a ich kvalitné počiatkové nastavenie. Z veľkého množstva dostupných metód v oblasti vývoja adaptívnych fuzzy systémov bol vybraný adaptívny neuro-fuzzy interferenčný systém Anfisedit (súčasť softvéru

prgramu MATLAB) [11]. Použilo sa subtraktívne zhľukovanie, ako aj metóda robustnej analýzy údajov s nasledujúcimi parametrami: *Range of influence* = 0.5, *Squash factor* = 1.1, *Accept ratio* = 0.45, and *Reject ratio* = 0.005.

Tabuľka 2. Namerané účinnosti turbíny v ustálených prevádzkových stavoch

Účinnosť η		ω_m				
		0,2	0,4	0,6	0,8	1
q	0,2	12,77%	30,36%	36,23%	37,83%	37,97%
	0,4	2,74%	26,40%	49,60%	64,41%	73,24%
	0,6	0,35%	13,64%	40,90%	66,92%	87,22%
	0,8	0,04%	5,68%	27,20%	56,26%	84,31%
	1	0,00%	2,11%	16,07%	42,03%	72,49%

Ako výsledok sme získali statický fuzzy systém typu Sugeno s dvanástimi pravidlami, ako je to znázornené na Obr. 8.

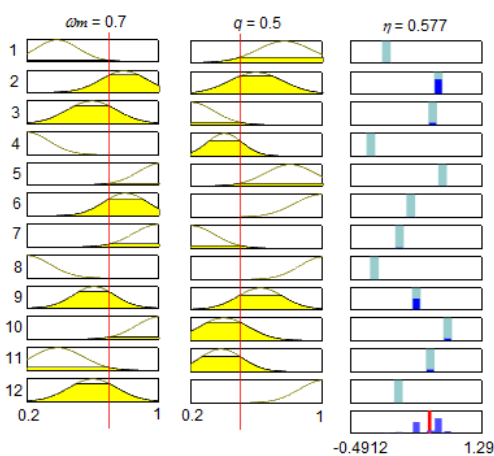
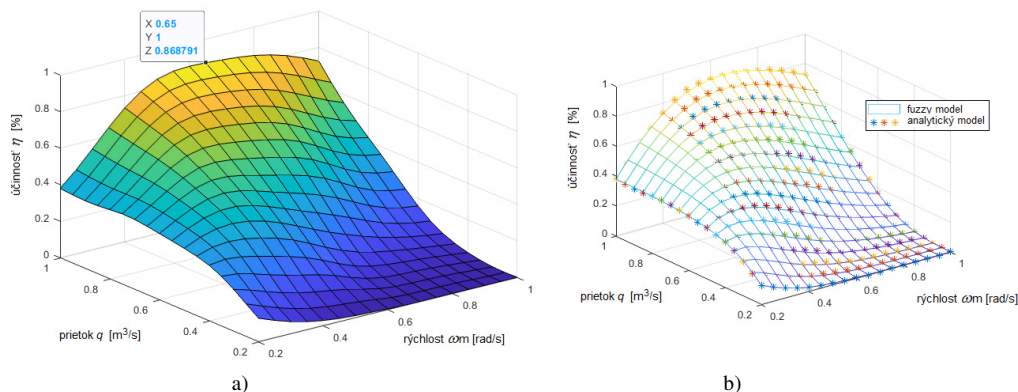


Fig.8 Štruktúra fuzzy modelu účinnosti hydraulickéj turbíny – Sugeno typ s 12 pravidlami

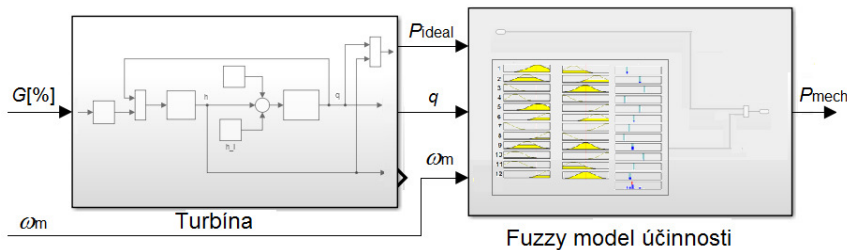
Navrhovaný fuzzy model nahrádza numerický výpočet účinnosti turbíny na základe rovníc (5)-(7). Tieto rovnice, ako už bolo spomenuté, obsahujú niektoré parametre, ktoré nevieme celkom presne určiť.

Z Obr.9 a) vidíme, že účinnosť na základe fuzzy modelu pre $\omega_m=0.65$ a $q=1$ je $\eta=87,62\%$. Na základe porovnania účinnosti získanej z fuzzy modelu a účinnosti vypočítanej analyticky (pozri Obr.7), môžeme konštatovať, že navrhnutý fuzzy model dosahuje takmer rovnakú presnosť ako analytický model. Ukazuje to aj grafické porovnanie účinnosti získaných analyticky a na základe fuzzy modelu, ktoré je zobrazené na Obr.9 b).



Obr.9 a) Fuzzy model účinnosti hydraulickéj turbíny
 b) Porovnanie analytického a fuzzy modelu účinnosti turbíny

Bloková schéma modelu hydraulickéj turbíny s fuzzy modelom jej účinnosti je znázornená na Obr.10.



Obr.10 Bloková schéma modelu hydraulickej turbíny so zohľadnením jej účinnosti

VII. ZÁVER

Predložený článok popisuje návrh koncepcie emulátora energetických systémov, ktorý má slúžiť pre výskum, simuláciu, verifikáciu konkrétnych energetických zariadení. Tento emulátor môže slúžiť aj ako tréningový nástroj pre školenie obsluhy takýchto systémov, pretože ich zaškoľovanie (predovšetkým v počiatočných fázach) na reálnom systéme je náročné a často hrozí nežiaducimi stratami, prípadne poškodením technologického zariadenia.

Emulátor podľa koncepcie popisovanej v článku je postavený v laboratóriách firmy EnergoControl s.r.o. Košice v spolupráci s Katedrou elektrotechniky a mechatroniky FEI TUKE.

Predložený článok popisuje aj stručný prehľad súčasných poznatkov a spôsobov v oblasti modelovania vodných turbín ako súčasti malých vodných elektrární. Je tu navrhnutý fuzzy model vodnej turbíny so zohľadnením jej účinnosti. Navrhnutý fuzzy model je možné získať z nameraných údajov prietoku vody turbínou a jej mechanickej uhlovej rýchlosti. Ukazuje sa, že navrhnutý fuzzy model dosiahol prakticky rovnakú presnosť ako analytické modely. Takýto fuzzy model je možné použiť v procese návrhu a vývoja optimálneho riadenia energie hydraulických turbín, ako aj v efektívnom procese určovania skutočnej miery opotrebovania turbíny.

POĎAKOVANIE

Článok bol publikovaný s podporou projektu APVV-16-0206 a APVV-19-0210.

LITERATÚRA

- [1] Y. Weijia, et al., "A Mathematical Model and Its Application for Hydro Power Units under Different Operating Conditions", *Energies*, vol. 8, 2015, pp. 10260-10275, DOI:10.3390/en80910260.
- [2] Ch.K. Weng, A. Ray, X.Dai, "Modelling of Power Plant Dynamics and Uncertainties for Robust Control Synthesis", *Applied Mathematical Modelling*, Vol. 20, Iss.7, 1996, pp. 501-512. DOI 10.1016/0307-904X(95)00169-K.
- [3] K.L. Koo, "Modeling of Plant Controllers Using the PSS-E to MATLAB Simulink Interface (PMSI) in PSS-E 30", *Power Technology*, Iss. 96, 2004, pp. 1-14.
- [4] A. Magnúsdóttir, D. Winkler, "Modelling of a Hydro Power Station in an Island Operation", *Proc. Of International Modelica Conference*, Prague, 2017, pp.483-492. DOI 10.3384/ecp17132483.
- [5] A.A. Usman, R.A. Abdulkadir, "Modelling and Simulation of Micro Hydro Power Plant Using MATLAB Simulink", *Proc. of 2nd Int. Conf. on Science, Technology and Management (ICSTM)*, University of Delphi, New Delphi, 2015, pp. 1121-1133.
- [6] M. Sattou, "Simulation Model of Hydro Power Plant Using MATLAB/Simulink", *Journal of Engineering Research and Applications*, Vol. 4, Issue 1, 2014, pp. 295-30.
- [7] J. Tiwari, et al., "Modelling and Simulation of Hydro Power Plant using MATLAB & WatPro 3.0", *I.J. Intelligent Systems and Applications*, Vol. 8, 2015, pp. 1-8. DOI: 10.5815/ijisa.2015.08.01.
- [8] A. Acakpovi, E.B. Hagan, F.X. Fiftatin, "Review of Hydropower Plant Models", *Int. Journal of Computer Applications*, Vol. 108, No. 18, 2014, pp. 33-38. DOI: 10.5120/19014-0541.
- [9] IEEE Working Group. Hydraulic turbine and turbine control models for system dynamic studies. *IEEE Trans on Power Syst.* 1992.
- [10] P. Fedor, D. Perdukova, "Model Based Fuzzy Control Applied to a Real Nonlinear Mechanical System", *Iranian Journal of Science and Technology, Transactions of Mechanical Engineering*, vol. 40, 2016, pp. 113-124. DOI 10.1007/S40997-016-0005-9.
- [11] D. Perduková, P. Fedor, M. Lacko, "DC Motor Fuzzy Model Based Optimal Controller", *MM Science Journal*, vol. 2021, 2021, pp. 4879-4885. DOI: 10.17973/MMSJ.2021_10_2021033.

Alternatívne metódy smerovania pre 6G siete

¹Gabriela VAŠKOVÁ, ²Natalia KURKINA, ³Ján PAPAJ, ⁴Maroš BAUMGARTNER

^{1,2,3,4} Katedra elektroniky a multimediálnych technológií, Fakulta elektrotechniky a informatiky, Technická Univerzita v Košiciach, Slovenská republika

¹gabriela.vaskova@student.tuke.sk, ²natalia.kurkina@tuke.sk, ³jan.papaj@tuke.sk, ⁴maros.baumgartner@tuke.sk

Abstract — Myšlienka vývoja siete 6 generácie v súčasnej dobe je veľmi populárna, ale zatiaľ 6G siete predstavujú len súbor teórií o tom, ako by mohli vyzeráť. Prezentovaný článok sa zameriava na preskúmanie možnosti použitia alternatívnych metód smerovania pre siete šiestej generácie. Alternatívne metódy na smerovanie, ktoré sú využité v tomto článku, sú metódy strojového učenia, ktoré sú podkategóriou Umelej inteligencie. V článku sú reprezentovaná mobilná ad hoc sieť, ktorá je súčasťou 6G siete na smerovanie, v ktorej sú využité metódy strojového učenia – Lineárna regresia a K-cluster. Sieť a proces smerovania sú simulované v programe MATLAB. Dosiiahnuté výsledky sa porovnávajú s klasickým princípom smerovania, keď sa používa najkratšia cesta od zdroja do cieľa.

Keywords — siete šiestej generácie, strojové učenie, umelá inteligencia, mobilná ad hoc sieť, smerovanie

I. ÚVOD

Siete šiestej generácie tvoria koncept budúcnosti v rozvíjaní bezdrôtových technológií. Avšak stále neexistujú. Vo výskumných prácach sú 6G siete popisované ako plne integrovaný internetový systém, ktorý umožňuje okamžitú komunikáciu medzi používateľmi, zariadeniami a okolitým prostredím [1]. Preto siete šiestej generácie prinášajú so sebou pokrok, ktorý sa dotýka predovšetkým sietí a komunikácie. Spôsobia revolúciu v týchto oblastiach tým, že budú viacvrstvové, nepredvídateľné, heterogénne a budú k dispozícii na veľkej škále. Výzvy, ktorým musia siete šiestej generácie čeliť sú bezproblémové pripojenie, stopercentná prevádzka, zaistená kvalita služieb, požiadavky na veľký počet zariadení, ktoré zahŕňajú aj neustále sa zväčšujúci počet IoT zariadení a veľký objem dát, ktoré pochádzajú z takýchto zariadení, ktoré je potrebné neustále spracovávať, analyzovať a prenášať [4].

Všetky tieto výzvy a problémy pri vývine 6G sietí sa dajú vyriešiť integráciou Umelej inteligencie do sietí šiestej generácie. Umelá inteligencia má vysokú schopnosť učenia sa, analýzy, klasifikácie a rozpoznávanie schopnosti. To všetko pomôže 6G sieťam implementovať optimalizáciu výkonu, získavanie vedomosti, sofistikované učenie, organizáciu štruktúry a komplikované rozhodovanie [4].

II. SMEROVANIE PRE SIETE ŠIESTEJ GENERÁCIE

Smerovanie (routing) je proces, v ktorom sa vyberá cesta na preposielanie dát cez jednu alebo viac sietí. Princípy smerovania sa môžu uplatniť na akýkoľvek typ siete, od telefónnych sietí až po verejnú dopravu. V sieťach sa vyberá trasa prípadne uzly pre smerovanie dát od ich zdroja k cieľu [3].

V tradičných sieťach je smerovač (router) sieťový hardvér zodpovedný za preposielanie paketov na miesta kam majú byť doručené. Smerovače sa pripájajú k dvom alebo viacerým sieťam a podľa potreby medzi nimi prenášajú dátové pakety. Používajú sa v domácnostiach a kanceláriách na nastavenie pripojenia k miestnej sieti. Výkonnejšie smerovače pracujú po celom internete a pomáhajú dátovým paketom dostať sa do cieľa [3].

V týchto sieťach je protokol spôsob formátovania údajov, aby každý pripojený počítač mohol údajom porozumieť. Smerovací protokol je protokol, ktorý sa používa na identifikáciu alebo oznamovanie ciest v sieti [3].

6G má víziu celých spektier a globálneho pokrytia, teda siete vesmír-vzduch-zem-more. Úplne zastúpenie aplikácií dokáže 6G dosiahnuť využitím Umelej inteligencie, strojového učenia a big data technológií.

6G tiež budú obsahovať aj rôzne typy komunikácie, ako napríklad komunikácia typu zariadenie-zariadenie, a hlavnou súčasťou takých sietí bude mobilná ad hoc sieť (MANET).

Mobilná ad hoc sieť je skupina bezdrôtových zariadení, ktoré navzájom komunikujú prostredníctvom bezdrôtových spojení tvoriacich bezdrôtovú sieť. Každý uzol v takej sieti je zároveň smerovač aj koncová stanica. Táto sieť patrí medzi siete bez infraštruktúry, ktoré nemajú pevnú topológiu. Vďaka tejto vlastnosti môžu byť nasadené v núdzových prípadoch, ako napríklad pri obnove po haváriách [7].

V 6G sieťach sa štandardné smerovacie protokoly, napríklad AODV, už nebudú dať použiť, práve z dôvodu, že v sieti sa bude nachádzať veľké množstvo terminálov. Tieto terminály sú mobilné, rýchlo sa pohybujú a prenášajú veľké množstvo dát. Preto je otázka optimálneho smerovania v takých sieťach dôležitá. Avšak, 6G siete sú smart siete alebo aj smart inteligentné siete a všetko v týchto sieťach rieši Umeľá inteligencia.

III. STROJOVÉ UČENIE A SIETE ŠIESTEJ GENERÁCIE

Strojové učenie (Machine Learning – ML) je silný nástroj Umelej inteligencie (Artificial intelligence – AI), obr. 1. Strojové učenie poskytuje systémom schopnosť, aby sa mohli automaticky učiť a zlepšovať zo skúsenosti, a to bez toho, aby boli vopred naprogramované. Zameriava sa na vývoj počítačových programov, ktoré majú prístup k údajom, aby sa mohli využiť na vlastné vzdelávanie. Proces učenia sa začína rôznymi pozorovaniami a údajmi. Údaje môžu byť príklady, skúsenosti alebo inštrukcie, vďaka ktorým je možné nájsť neskôr vzory v údajoch. Hlavným cieľom je umožniť počítačom automatické učenie bez nutnosti zásahu človeka alebo jeho asistencie a podľa toho vykonať akciu [1].



Obr. 1 Strojové učenie

Z výskumov vyplýva, že strojové učenie a 6G sú vzájomne výhodné. Na jednej strane 6G môže podporovať služby strojového učenia. Na druhej strane strojové učenie bude optimalizovať architektúru 6G a bude hrať dôležitú úlohu v štruktúre 6G [5].

Charakteristiky strojového učenia sú [5]:

1. Centralizácia - Väčšina aplikácií strojového učenia je centralizovaných, napríklad použité údaje monopolizujú konkrétne spoločnosti.
2. Neprehľadnosť - V tejto fáze je proces strojového učenia skôr ako riešenie čiernej skrinky. Ľudia môžu ovládať iba výstup, ale nedokážu úplne pochopiť dôvody od prechodového procesu po výstup.
3. Vysoká účinnosť - Výpočtový výkon strojového učenia je veľmi silný. S vývojom algoritmu a podporou 6G bude strojové učenie efektívnejšie.

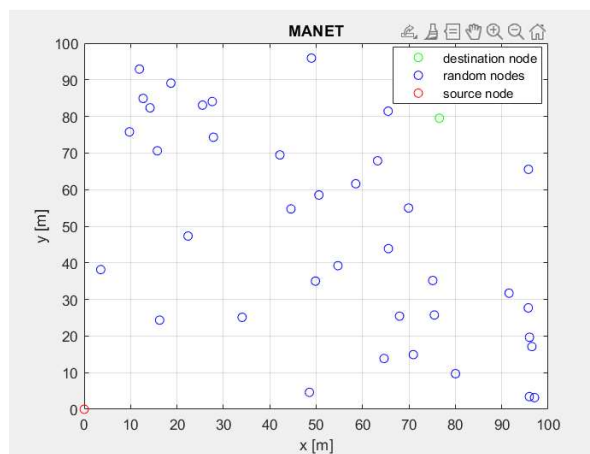
Hlavné dôvody používania strojového učenia sú [6]:

- Vyriešiť problém – Problémov, ktoré dokáže strojové učenie vyriešiť je mnoho, napríklad obchodné problémy, marketing, zisk, skrátenie času alebo viac výroby.
- Technický úspech – Mnoho spoločností súťaží o úspechy, ktoré dosiahnu pomocou strojového učenia.

Pri riešení súčasných problémov sa používajú techniky strojového učenia ako sú učenie s učiteľom, učenie bez učiteľa a učenie s posilňovaním [4]. Tieto techniky sa používajú pri riešení problémov klasifikácie, regresie, zhukovania, asociácie, predpovedania atď. Metóda klasifikácie strojového učenia je vhodnou metódou na predpovedanie sieťového prenosu a určenie najpravdepodobnejších uzlov, ktoré sa môžu v danom smere vyskytnúť, a ktoré je možné použiť na smerovanie v sieťach šiestej generácie.

IV. SIMULÁCIA A ANALYZOVANIE VÝSLEDKOV

Na vytvorenie programu simulovania siete a procesu smerovania bol použitý nástroj MATLAB. Program bol zostavený tak že veľkosť plochy, na ktorej je sieť rozmiestnená, počet mobilných uzlov a rádiový dosah uzlov, v ktorom ak sa budú uzly nachádzať môžu medzi sebou navzájom komunikovať, si zvolí používateľ pri spustení. Zdrojový a cieľový uzol sa generuje náhodne. Pre každý vygenerovaný uzol sú vygenerované parametre - bandwidth, data rate a latency. Po vykreslení pôvodnej polohy uzlov nasleduje zápis polohy a parametrov do textového súboru, ktoré budú vstupovať do strojového učenia. Ďalej nasleduje výpočet vzdialenosti medzi jednotlivými uzlami, z čoho sa potom vypočíta, či sa jednotlivé uzly nachádzajú v rádiovom dosahu a môžu navzájom komunikovať alebo nie.



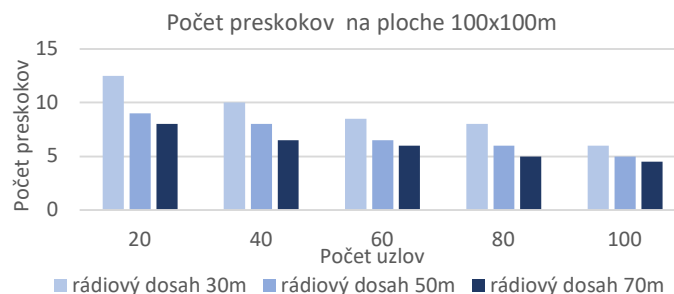
Obr. 2 Vytvorená mobilná ad hoc sieť

V prípade, že sa v okolí zdrojového a cieľového uzla nachádzajú uzly, ktoré sú v ich rádiovom dosahu nasleduje nájdenie a vykreslenie najkratšej trasy medzi nimi. V prípade, ak zdrojový a cieľový uzol nemajú v svojom rádiovom dosahu žiadne uzly, vypíše sa: „Cesta nie je možná.“

Na vyber vhodných uzlov boli použité dva algoritmy strojového učenia – lineárna regresia a k-cluster. Úlohou bolo aplikovať strojové učenie do procesu smerovania. Keďže v 6G sieťach sa štandardne protokoly na smerovanie už nebudú dať použiť, z dôvodu, že v sieťach sa bude nachádzať veľa zariadení. Rovnako ani pri strojovom učení sa už nebudú dať použiť štandardne smerovacie prístupy ako boli doteraz. Už sa nehľadá cesta z bodu A do bodu B, ale vyberajú sa uzly, ktoré majú preniesť danú správu. Z dát ako je poloha, šírka pásma, dátová rýchlosť a oneskorenie bol v prave vytvorený dataset, na ktorom sa vybrané algoritmy učili. Vyberali sa vhodné uzly na smerovanie. Vybrané algoritmy vyberajú cestu nie len z dôvodu, že je cesta najkratšia, ale aj z pohľadu, ktoré uzly majú nízke oneskorenie, vysokú dátovú rýchlosť a vyššiu šírku pásma.

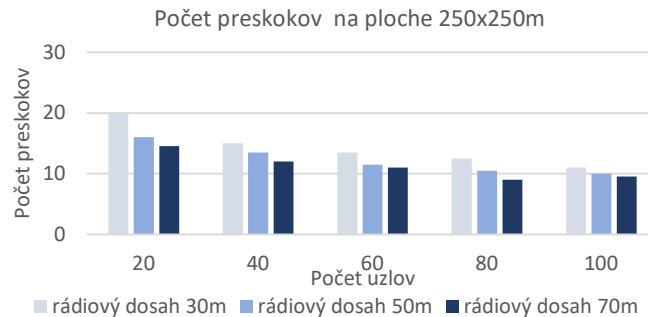
V MATLABE bol skúmaný priemerný počet preskokov na plochách o veľkosti 100x100m, 250x250m a 500x500m, pričom skúmaný počet uzlov sa menil od 20 po 100 a to konkrétne 20, 40, 60, 80 a 100. Rádiový dosah uzlov sa menil od 30 cez 50 až po 70m.

Napríklad na ploche 100x100m pri počte uzlov 60 a rádiovom dosahu 50m bol priemerný počet preskokov 7.5, obr. 3.



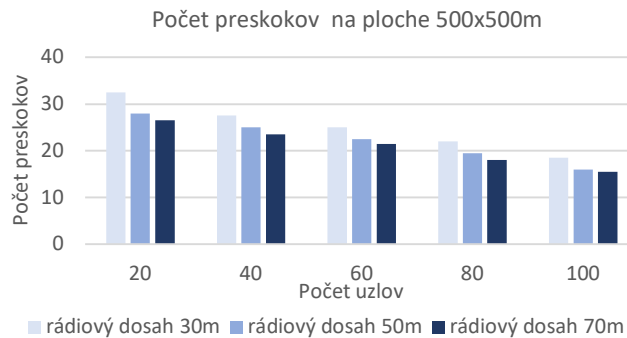
Obr. 3 Počet preskokov na ploche 100x100

Na ploche 250x250m pri počte uzlov 20 a rádiovom dosahu 30m bol priemerný počet preskokov 20, obr. 4.



Obr. 4 Počet preskokov na ploche 250x250

Na ploche 500x500m pri počte uzlov 80 a rádiovom dosahu 70m bol priemerný počet preskokov 18.5, obr. 5.



Obr. 5 Počet preskokov na ploche 500x500

V. ZÁVER

V článku je stručne popísane čo je sieť šiestej generácie a možnosti využitia strojového učenia pri smerovaní v takýchto sieťach.

Sieť MANET je hlavnou súčasťou siete šiestej generácie. Preto práca tiež obsahuje aj vytvorenú mobilnú ad hoc sieť, v ktorej spolu môžu komunikovať nie len mobilné zariadenia, ale napríklad aj senzory. Uzly na prenos sa vyberali nie len pomocou tejto mobilnej ad hoc siete, ale tiež pomocou algoritmov strojového učenia a to konkrétne pomocou algoritmov Linear regression a K-cluster. Z dosiahnutých výsledkov vyšiel ako najlepší algoritmus strojového učenia Linear regression, pretože vyberal uzly s väčšou dátovou rýchlosťou a menším oneskorením na rozdiel od ostatných algoritmov.

POĎAKOVANIE

Táto publikácia vznikla vďaka podpore v rámci Operačného programu Integrovaná infraštruktúra pre projekt: Výskum v sieti SANET a možnosti jej ďalšieho využitia a rozvoja, kód ITMS: NFP313010W988, spolufinancovaný zo zdrojov Európskeho fondu regionálneho rozvoja.

REFERENCES

- [1] Tang, F., Kawamoto Y., Kato, N., Liu, J.: Future Intelligent and Secure Vehicular Network Toward 6G: Machine-Learning Approaches. 2019.
- [2] Taeho Jo: Machine Learning Foundations: Supervised, Unsupervised, and Advanced Learning. 2021.
- [3] Deep Medhi, Karthik Ramasamy: Network Routing Algorithms, Protocols, and Architectures. 2018.
- [4] Sheth, K., Patel, K., Shah, H., Tanwar, S., Gupta, R., Kumar, N.: A taxonomy of AI techniques for 6G communication networks. 2020.
- [5] Sun, Y., Liu, J., Wang, J., Cao, Y., Kato, N.: When Machine Learning Meets Privacy in 6G: A Survey. 2020-07-26.
- [6] Asadi, S.: Space Delay-Tolerant Networks Routing Using Artificial Neural Networks. 2018.
- [7] Sankaran, K., Devabalaji, K., Babu, T., Alhelou, H., Tuvaraj, T. A Recurrent Reward Based Learning Technique for Secure Neighbor Selection in Mobile AD-HOC Networks. 2021-01-21
- [8] Le, D., Pandez, A. K., Tadepalli, S., Rathore, P. S., Chatterjee, J. M.: Network Modeling, Simulation and Analysis in MATLAB. 2019.

Napät'ový regulátor pre nepriamy jednosmerný menič s mäkkým spínaním

¹Marek PÁSTOR, ²Jaroslava ŽILKOVÁ

^{1,2}Katedra elektrotechniky a mechatroniky, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Košice, Slovenská republika

¹marek.pastor@tuke.sk, ²jaroslava.zilkova@tuke.sk

Abstrakt — Článok sa zaoberá analýzou a návrhom napät'ového PI regulátora pre riadenie výstupného napätia nepriameho jednosmerného meniča s mäkkým spínaním. Dynamika meniča je aproximovaná spojítym modelom druhého rádu. Na základe tohto modelu je navrhnutý PI regulátor na základe požadovanej dynamiky výstupného napätia meniča. Spojitý PI regulátor je následne diskretizovaný a implementovaný pomocou mikotroktroléra STM 32F746. Činnosť regulátora je overená meraním na 2 kW laboratórnom modeli meniča.

Kľúčové slová — jednosmerný menič, mäkké spínanie, PI regulátor

I. ÚVOD

Jednosmerné meniče predstavujú kľúčový komponent v topológiách meničov pre obnoviteľné zdroje energie a elektromobilitu. Moderné elektrické vozidlá, obnoviteľné zdroje energie a napájacie zdroje vyžadujú meniče s vysokou účinnosťou, nízkou hmotnosťou a vysokou spoľahlivosťou. V mobilných aplikáciách zohráva hmotnosť a rozmery meniča veľmi dôležitú úlohu. To so sebou prináša neustálu požiadavku na zvyšovanie spínacej frekvencie meniča. Nevýhodou tohto prístupu je zvyšovanie spínacích strát, čo so sebou prináša zvýšené požiadavky na chladenie meniča [1]-[8]. Na zníženie spínacích strát meniča boli vytvorené rôzne techniky mäkkého spínania nepriameho meniča v mostíkovom zapojení [2]-[7]. Zvyčajne sa využíva aktívny odľahčovací obvod [2] alebo riadený usmerňovač [3]-[7].

Pre riadenie výstupného napätia jednosmerného meniča sa zvyčajne používa napät'ové alebo prúdové riadenie [9]. Tradičným riešením je PID regulátor so šírkovo impulzovou moduláciou [9]-[18]. Dôvodom je jednoduchosť PID regulátora a jeho rozšírenie v priemyselnej praxi.

Mnohé výkonové polovodičové meniče obsahujú nelinearity spôsobené použitými súčiastkami alebo spôsobom riadenia. Klasický PID regulátor má problém pri riadení nelineárnych systémov. Preto bolo vytvorených množstvo moderných riadiacich metód, ktoré sú založené na modeli meniča alebo na báze znalostí. Riadenia ako je prediktívne riadenie sa stali populárnymi s nárastom výpočtového výkonu mikrokotrolérov [19]-[24]. Tie sa využívajú hlavne v meničoch pripojených do siete [20]-[22] a v elektrických pohonoch [23], [24]. Spomedzi metód založených na báze pravidiel je najčastejšie fuzzy riadenie [25]-[31].

Umelé neurónové siete sa v posledných rokoch uplatňujú aj v oblasti výkonovej elektroniky [31]-[34]. Umožňujú identifikáciu zát'aže a optimalizáciu parametrov PID regulátora [33]. Taktiež existujú aplikácie genetických algoritmov v oblasti výkonovej elektroniky [36]-[37].

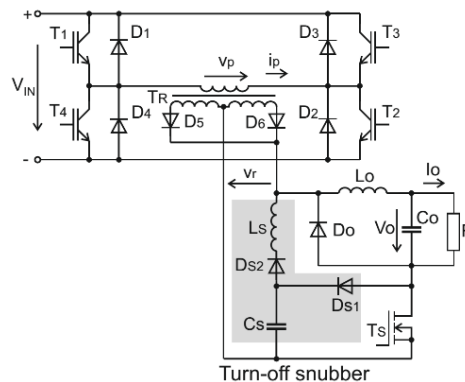
Článok opisuje aplikáciu PI regulátora pre riadenie výstupného napätia nepriameho meniča v mostíkovom zapojení s odľahčovacím obvodom na sekundárnej strane (Obr. 1). model meniča je aproximovaný sústavou druhého rádu pre ktorú je navrhnutý PI regulátor. Činnosť regulátora je overená meraním na laboratórnom modeli meniča.

II. NEPRIAMY JEDNOSMERNÝ MENIČ

A. Topológia meniča a princíp činnosti

Na Obr. 1 je znázornený nepriamy jednosmerný menič v mostíkovom zapojení s mäkkým spínaním. Menič pozostáva z mostíkového striedaa tvoreného IGBT tranzistorami $T_1 - T_4$ s antiparalelnými diódami $D_1 - D_4$. Striedač napája vysokofrekvenčný transformátor TR. Výstup

transformátora je dvojcestne usmernený riadeným usmerňovačom tvoreným diódami D5 a D6 a tranzistorom TS a vyhladený výstupným filtrom (L_o , C_o). Vypínacie straty tranzistora T_s sú odstránené použitím odľahčovacieho obvodu pozostávajúceho z kondenzátora C_s , diód D_{s1} a D_{s2} a cievky L_s . Spínacie straty striedača sú odstránené použitím riadeného usmerňovača, ktorý vypína skôr ako tranzistory striedača. Tie tak vypínajú len magnetizačný prúd transformátora. Výstupné kapacity tranzistorov striedača sú vybité magnetizačným prúdom transformátora počas trvania bezpečnostnej medzery, čím sa zabezpečí zapínanie tranzistorov v nule napätia. Vypínacie straty tranzistora T_s sú minimalizované použitím odľahčovacieho obvodu s rekuperáciou energie. Parametre meniča sú uvedené v tabuľke 1.



Obr. 1 Schéma zapojenia meniča

Tabuľka 1
Základné parametre meniča

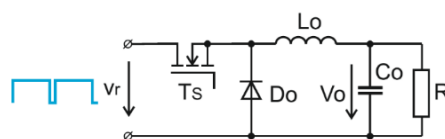
Parameter	Hodnota	Jednotka
Výstupný výkon P_{OUT}	2	kW
Napájacie napätie V_{IN}	300	V
Spínacia frekvencia f_s $T_1 - T_4$	50	kHz
Nominálne výstupné napätie V_o	40	V
Nominálny výstupný prúd I_o	50	A
Prevod transformátora p T_R	5	-
Kapacita výstupného filtra C_o	27	μF
Indukčnosť výstupného filtra L_o	4	μH
Kapacita odľahčovacieho obvodu C_s	220	nF
Indukčnosť odľahčovacieho obvodu L_s	300	nH

B. Model jednosmerného meniča

Pre účely návrhu regulátora bol nepriamy jednosmerný menič aproximovaný znižujúcim meničom napájaným usmerneným napätím v_r ako je znázornené na Obr. 2. Prenosová funkcia meniča je aproximovaná sústavou druhého rádu (1). Toto zjednodušenie zanedbáva zvlhnie spôsobené spínaním meniča, no umožňuje výrazne zjednodušiť model meniča.

$$P(s) = \frac{v_o(s)}{d_{T_s}(s)} = \frac{K_C \omega_n^2}{s^2 + 2\zeta \omega_n s + \omega_n^2} \quad (1)$$

Prenosová funkcia obsahuje tri parametre, ktoré potrebujú byť identifikované. Zosilnenie K_C je úmerné strednej hodnote výstupného napätia meniča a zahŕňa zosilnenie modulátora. Uhlová frekvencia systému ω_n je hlavne ovplyvnená rezonančnou frekvenciou LC filtra. Tlmenie ζ je ovplyvnené odporom záťaže a použitou striedou spínania usmerňovacieho tranzistora.

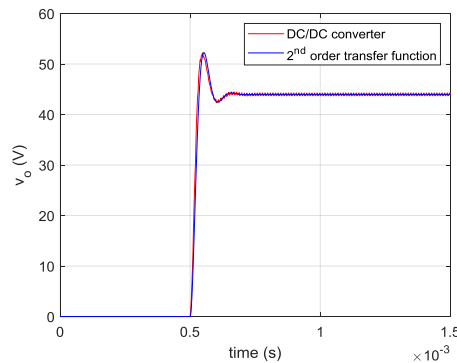


Obr. 2 Aproximácia nepriameho meniča znižovací meničom

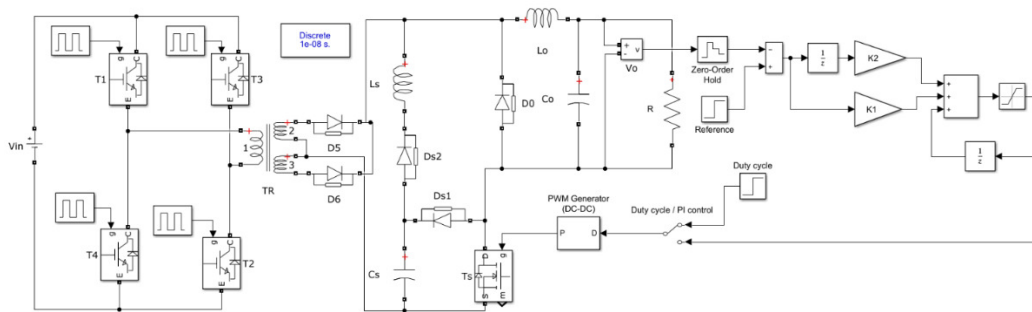
Na Obr. 3 je znázornené porovnanie aproximovanej prenosovej funkcie meniča a prechodová funkcia získaná simuláciou meniča v prostredí Matlab/Simulink s využitím knižnice Simscape

(Obr. 4). Prenosová funkcia je výrazne ovplyvnená striedou spínania tranzistora T_s . Pre aproximáciu prenosu meniča pri striede 0,7 bol menič aproximovaný sústavou druhého rádu opísanou prenosovou funkciou (2).

$$P(s) = \frac{2 \cdot 10^{11}}{s^2 + 6,33 \cdot 10^4 s + 4,54 \cdot 10^9} \quad (2)$$



Obr. 3 Porovnanie prechodovej funkcie jednosmerného meniča s prechodovou funkciou získanou aproximáciou sústavou druhého rádu



Obr. 4 Programová schéma pre simuláciu meniča v prostredí Simulinku

III. PI REGULÁTOR

A. Návrh PI regulátora

Prenosová funkcia (2) bola použitá pre návrh PI regulátora. Menič predstavuje stabilnú sústavu, takže hlavnou úlohou regulátora je zabezpečiť reguláciu výstupného napätia meniča. PI regulátor je v spojitej oblasti navrhnutý s ohľadom na požadovanú časovú konštantu uzavretého regulačného obvodu [13].

$$C(s) = \frac{U(s)}{E(s)} = K_p + \frac{K_I}{s} \quad (3)$$

Prenosová funkcia uzavretého obvodu s PI regulátorom je:

$$F(s) = \frac{K_C K_p \omega_n^2 s + K_C K_I \omega_n^2}{s^3 + 2\zeta \omega_n s^2 + (K_C K_p + 1) \omega_n^2 s + K_C K_I \omega_n^2} \quad (4)$$

Na základe [13] sú póly prenosu (4) umiestené podľa (5).

$$s^3 + 2\zeta \omega_n s^2 + (K_C K_p + 1) \omega_n^2 s + K_C K_I \omega_n^2 = \left(s + \frac{1}{\tau}\right)^3 \quad (5)$$

kde τ je dominantná časová konštantu uzavretého systému. Koeficienty PI regulátora sú tak:

$$K_p = \frac{3/\tau^2 - \omega_n^2}{K_C \omega_n^2}; \quad K_I = \frac{1}{K_C \omega_n^2 \tau^3} \quad (6)$$

Regulátor bol implementovaný v digitálnej forme pomocou mikrokontroléra STM32F746. Jeho spojitý tvar tak bolo potrebné diskretizovať tak, že akčný zásah je definovaný rovnicou:

$$u[k] = u[k-1] + K_I e[k] + K_2 e[k-1] \quad (7)$$

Kde zosilnenia K_1 a K_2 sú získané diskretizáciou lichobežníkom (Tustinova aproximácia):

$$\begin{cases} K_1 = K_p + K_I T_{sample} / 2 \\ K_2 = -K_p + K_I T_{sample} / 2 \end{cases} \quad (8)$$

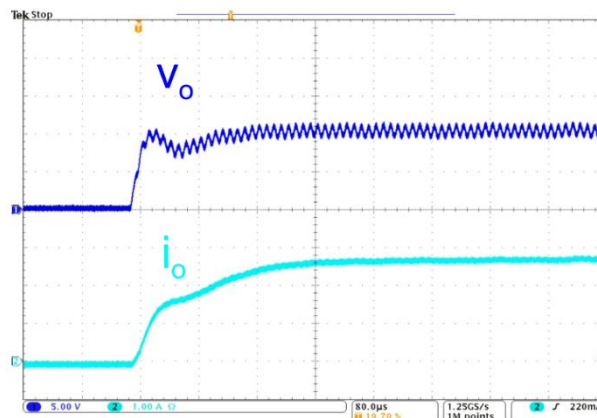
Vzorkovací čas T_{sample} v (8) je nastavený podľa spínacej frekvencie usmerňovacieho tranzistora T_s , čo predstavuje dvojnásobok spínacej frekvencie primárnych tranzistorov f_s .

Tabuľka 2
 Základné parametre meniča

Parameter	Hodnota	Jednotka
Vzorkovacia perióda regulátora T_{sample}	10	μs
Dominantná časová konštanta τ	35	μs
Zosilnenie K_1	0,0095	-
Zosilnenie K_2	-0,007	-

B. Verifikácia PI regulátora

Činnosť navrhnutého diskretného PI regulátora bola overená meraním na laboratórnom modeli meniča. Želaná hodnota výstupného napätia meniča bola skokovo zmenená na 10 V. Doba ustálenia výstupu meniča je približne 120 μs . Výstup meniča tak dosiahol požadovanú hodnotu v rámci päťnásobku dominantnej časovej konštanty sústavy. Výstupné napätie nemá prekmit a výstup meniča dosiahol očakávanú dynamiku.



Obr. 5 Odmeraná prechodová charakteristika výstupného napätia a prúdu meniča

IV. ZÁVER

Článok opisuje návrh diskretného regulátora pre nepriamy menič s mäkkým spínaním. Návrh regulátora je založený na aproximácii meniča sústavou druhého rádu a následnej diskretizácii navrhnutého spojitého regulátora. Činnosť regulátora bola overená meraním na laboratórnom modeli meniča. Výstup meniča dosiahol požadovanú dynamiku a navrhnutý PI regulátor pracoval podľa očakávaní. V ďalšom kroku bude diskretný PI regulátor porovnaný s fuzzy regulátorom, ktorý by mal vzhľadom na nelinearitu meniča dosahovať lepšie výsledky.

REFERENCIE

- [1] G. Kácsor, P. Špánik, J. Dudrik, M. Luft, and E. Szychta, "Principles of operation of three-level phase shift controlled converter", *Electronics and Electrical Engineering*, vol. 82, no. 2, pp. 69-74, 2008.
- [2] M. Bodor, J. Dudrik and J. Perdulak, "ZVZCS PWM converter using secondary active clamp", *Acta Electrotechnica et Informatica*, vol. 11, no. 3, pp. 26–30, 2011, [online] Available: <http://dx.doi.org/10.2478/v10198-011-0026-7>.
- [3] J. Dudrik and V. Ruscin, "Voltage fed zero-voltage zero-current switching PWM DC-DC converter," 13th International Power Electronics and Motion Control Conference, (*EPE-PEMC*) pp. 295-300, 2008, [online] Available: doi:10.1109/EPEPEMC.2008.4635281.
- [4] J. Dudrik, M. Pástor, M. Lacko and R. Žatkovič, "High-frequency soft-switching PWM DC-DC converter with active output rectifier operating as a current source for arc welding applications," *Electric Power Components and Systems*, vol. 45, no. 6, pp. 681-691, 2017.
- [5] J. Dudrik, M. Pástor, M. Lacko and R. Žatkovič, "Zero-voltage and zero-current switching PWM DC-DC converter using controlled secondary rectifier with one active switch and nondissipative turn-off snubber," *IEEE Transactions on Power Electronics*, vol. 33, no. 7, pp. 6012-6023, 2018.
- [6] J. Dudrik ... [et al.], "ZVZCS PWM DC-DC converter with controlled secondary rectifier for arc welding," *Acta Electrotechnica et Informatica*. vol. 16, no. 2, pp. 33-41, 2016,.: [online] Available: DOI: 10.15546/aei-2016-0013

- [7] J. Dudrik, V. Ruščin, "ZVZCS PWM DC-DC converter with controlled output rectifier," *Acta Electrotechnica et Informatica*, vol. 10, no. 1, pp. 12-17, 2010.
- [8] H. M. Rashid, *Power Electronics Handbook*, ISBN 978-0-12-382036-5, pp. 1409, 2011.
- [9] M. Leso, J. Zilková, J., M. Biroš and P. Talian, "Survey of control methods for DC-DC converters," *Acta Electrotechnica et Informatica*, vol. 18, no. 3, pp. 41-46, 2018, [online] Available: DOI: 10.15546/aei-2018-0024.
- [10] S. Verma, S.K. Singh and A.G. Rao, "Overview of control techniques for DC-DC converters", *Research Journal of Engineering Sciences*, vol.2, no. 8, pp. 18-21, 2013, ISSN 2278 – 9472.
- [11] A.I. Bogdan and N. Bizon, "Voltage-mode control of the DC-DC power converter - a short review", 7th International Conference on Electronics, Computers and Artificial Intelligence, pp.E-27-E-32, 2015.
- [12] Jian Sun: *Dynamics and control of switched electronic systems, Chapter 2: Pulse-Width Modulation*, ISBN: 978-1-4471-2884-7, 2012.
- [13] F. Deniz, B. Alagoz, and N. Tan, "PID controller design based on second order model approximation by using stability boundary locus fitting," *Proc. 9th Int. Conf. Elect. Electron. Eng.*, pp. 827–831, Nov. 2015.
- [14] N. George, V. N. Panchalai, E. Sebastian and S. Narayanan, "Digital voltage-mode-control of a full-bridge phase-shift-modulated DC-DC converter," *Annual International Conference on Emerging Research Areas: Magnetics, Machines and Drives (AICERA/iCMMD)*, pp. 1-6, 2014.
- [15] G. Malleshm, S. Mishra, A.N. Jha, "Maiden application of Ziegler-Nichols method to AGC of distributed generation system," *Power Systems Conference and Exposition, PSCE '09 IEEE/PES*, pp.1-7, 2009.
- [16] H. Jingqing, "From PID to active disturbance rejection control," *IEEE Transactions on Industrial Electronics*, vol.56, no.3, pp.900-906, March 2009.
- [17] M. Namnabat, M. B. Poodeh and S. Eshtehardiha, "Comparison of the control methods in improvement the performance of the DC-DC converter", *ICPE '07, 7th International Conference on Power Electronics*, pp.246-251, 2007.
- [18] M. P. Kazmierkowski, F. Blaabjerg and R. Krishnan, *Control in power electronics-selected problems*. Elsevier Science, San Diego, 2002.
- [19] S. Vazquez, J.I. Leon, L.G. Franquelo, J. Rodriguez, H.A. Young, A. Marquez, P. Zanchetta, "Model predictive control: A review of its applications in power electronics", *Industrial Electronics Magazine, IEEE*, vol.8, no.1, pp.16-31, March 2014.
- [20] M. Pastor, J. Dudrik and M. Vacek, "Predictive power control of grid-tied multilevel inverter", *Acta Electrotechnica et Informatica*, vol. 13, no.3, pp. 3–7, 2013, DOI: 10.2478/aei-2013-0031.
- [21] M. Pastor and J. Dudrik, "Grid-tied 15-level cascade inverter with predictive current control", *Electronics and Electrical Engineering*, vol. 18, no. 9 pp. 19-22, 2012, <http://www.eejournal.ktu.lt/index.php/elt/article/view/2798/1980>.
- [22] M. Pastor and J. Dudrik, "Predictive current control of grid-tied cascade H-bridge inverter", *Automatika, Journal for Control, Measurement, Electronics, Computing and Communications*, vol. 54, no. 3, pp. 308-315, 2013, <https://automatika.korema.hr/index.php/automatika/article/view/186/256>
- [23] K. Kyslan, V. Šlapák, F. Ďurovský, V. Fedák and S. Padmanaban, "Feedforward finite control set model predictive position control of PMSM," 2018 IEEE 18th International Power Electronics and Motion Control Conference (PEMC), pp. 549-555, 2018, doi: 10.1109/EPEPEMC.2018.8521989
- [24] P. Cortes, M. P. Kazmierkowski, R. M. Kennel, D. E. Quevedo and J. Rodriguez, "Predictive control in power electronics and drives", *IEEE Trans. Industrial Electronics*, vol. 55, no. 12, pp. 4312–4324, 2008. [Online]. Available: <http://dx.doi.org/10.1109/TIE.2008.2007480>
- [25] G. M. T. Nguyen, T.D.Mai, C.B.Tran, H.P.Nguyen, "PID-fuzzy logic hybrid controller for grid-connected photovoltaic inverters," *International Forum on Strategic Technology, (IFOST)*, pp.140-144, 13-15 Oct. 2010.
- [26] M. Louzani, E. Aroudam, "Control and stabilization of three-phase grid connected photovoltaics using PID-Fuzzy logic, IEEE International Conference on Intelligent Energy and Power Systems (IEPS), pp.279-284, 2-6 June 2014.
- [27] G. Liping; J.Y. Hung, R.M. Nelms, "Evaluation of DSP-based PID and fuzzy controllers for DC–DC converters, *IEEE Transactions on Industrial Electronics*, vol.56, no.6, pp.2237-2248, June 2009.
- [28] Jawhar S, J.; Marimuthu, N.S.; Singh N, A.: "A Neuro-Fuzzy Controller for a Non Linear Power Electronic Boost Converter, *International Conference on Information and Automation*, pp.394-397, 15-17 Dec. 2006.
- [29] M. Leso, J. Zilkova and P. Girovsky, "Development of a simple fuzzy logic controller for DC-DC converter," *PEMC 2018, 18th IEEE International Conference on Power Electronics and Motion Control*, pp. 86-93, 2018, DOI: [10.1109/EPEPEMC.2018.8521896](https://doi.org/10.1109/EPEPEMC.2018.8521896).
- [30] M. Leso, J. Zilkova, M. Pastor and J. Dudrik, "Fuzzy logic control of soft-switching DC-DC converter," *Elektronika ir Elektrotechnika*, vol. 22, no 5, pp.3-7, 2016, <https://doi.org/10.5755/j01.eie.22.5.16334>.
- [31] D. Perduková and P. Fedor, "Design rules of a fuzzy controller using genetic algorithms", *Proc. of the 7th Int. Scientific Symposium on Electrical Power Engineering, Elektroenergetika 2013*, pp. 513-516, 2013, ISBN 978-80-553-1441-9.
- [32] B. K. Bose, "Neural network applications in power electronics and motor drives-An introduction and perspective", *IEEE Transactions on Industrial Electronics*, vol.54, no.1, pp.14-33, Feb. 2007.
- [33] H. Maruta, M. Motomura and F. Kurokawa, "An evaluation study on circuit parameter conditions of neural network controlled DC-DC converter", *12th International Conference on Machine Learning and Applications, (ICMLA)*, vol.2, pp.249-254, 4-7 Dec. 2013.
- [34] Y. A. R. Mohamed and E. F. El-Saadany, "Adaptive discrete-time grid-voltage sensorless interfacing scheme for grid-connected DG-inverters based on neural-network identification and deadbeat current regulation", *IEEE Transactions on Power Electronics*, vol. 23, no. 1, pp. 308-321, 2008
- [35] P. Vas, *Artificial-Intelligence-Based Electrical Machines and Drives*, Oxford Press, New York, 1999.
- [36] K. S. Kostov, J. J. Kyyra, "Genetic algorithm optimization of peak current mode controlled buck converter", *Proceedings of the 2005 IEEE Mid-Summer Workshop on Soft Computing in Industrial Applications, SMCia/05*, pp.111-116, 28-30 June 2005.
- [37] Y. Nettari, M.N. Harmas, "Genetic algorithm based adaptive fuzzy terminal synergetic DC-DC converter control", *3rd International Conference on Control, Engineering & Information Technology (CEIT)*, pp.1-7, May 2015, *Proc. IEEE 18th International Power Electronics and Motion Control Conference (PEMC)*, 2018, pp. 86–93

Návrh a realizácia FTTH siete

¹Luboš OVSEŇÍK, ²Jakub ORAVEC, ³Norbert ZDRAVECKÝ, ⁴Maroš LAPČÁK,
⁵Samuel ANDREJČÍK

^{1,2,3,4,5} Katedra elektroniky a multimediálnych telekomunikácií, Fakulta elektrotechniky
a informatiky, Technická univerzita v Košiciach, Slovenská republika

¹lubos.ovsenik@tuke.sk, ²jakub.oravec@tuke.sk, ³norbert.zdravecky@tuke.sk,
⁴maros.lapcak@tuke.sk, ⁵samuel.andrejcik@tuke.sk

Abstrakt — Príspevok sa zberá problematikou pasívnych optických vláknových sietí (PON) na báze TDM, ich návrhom, simuláciou, ako aj ich fyzikálnym meraním pomocou OTDR metódy. Kládli sme si za cieľ podrobne analyzovať postup navrhovania a budovania optických prístupových sietí typu FTTH využívajúcich pasívny štandard GPON. Výsledkom je návrh a simulácie optickej FTTH siete Prešov-Nižná Šebastová.

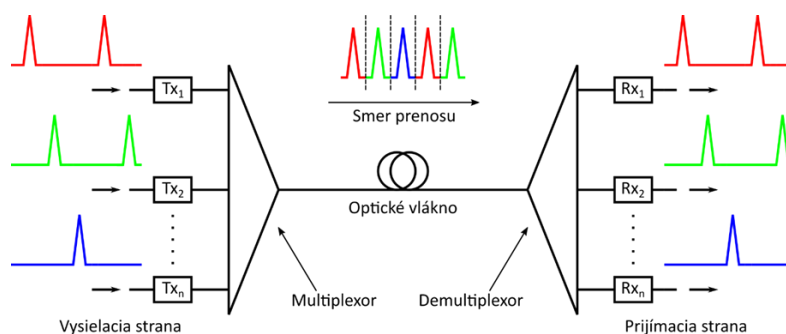
Kľúčové slová — FTTH, GPON, optické pasívne siete, PON, TDM

I. ÚVOD

Udalosti posledných rokov a prechod mnohých ľudí na prácu z domu zapríčinili výrazné zvýšenie dopytu po vysokorychlostnom pripojení. Poskytovatelia telekomunikačných služieb sú tak pre zachovanie konkurencieschopnosti nútení nahradzovať prístupové siete založené na zastaralých technológiách (metalických kábloch) za moderné optické prístupové siete. História súčasných optických telekomunikácií sa začína v 60. rokoch 20. storočia vynálezom lasera. Optické telekomunikácie boli v počiatku z dôvodu vysokých finančných nákladov a pomerne nízkej šírky pásma využívané len v chrbticových sieťach. Počas posledných desaťročí však technológie v oblasti optických komunikácií výrazne pokročili a dnes je možné v rámci jedného optického vlákna pomocou multiplexovania (TDM, WDM) vysielat' na frekvenciách dosahujúcich stovky terahertzov (THz). Významným míľnikom bol vynález optického rozbočovača – splittra, vďaka ktorému sa optické komunikácie začali aplikovať aj do prístupových sietí. Používaním optických rozbočovačov v prístupových sieťach je možné pripojiť až 128 koncových užívateľov na jedno optické vlákno, čím sa výrazne znižujú finančné náklady a zvyšuje efektívnosť siete. Motiváciou k tvorbe tejto práce boli rýchlo sa rozrastajúce optické prístupové siete FTTH, v ktorých nasadzovaní patrí Slovensko medzi popredné európske krajiny.

II. PASÍVNA OPTICKÁ SIEŤ NA BÁZE TDM

Multiplex s časovým delením TDM (Time Division Multiplex) predstavuje v oblasti digitálnych telekomunikačných sietí metódu prenosu informácií, ktorá spočíva v prenášaní minimálne dvoch alebo viacerých na sebe nezávislých digitálnych signálov prostredníctvom jedného zdieľaného širokopásmového komunikačného kanála. V oblasti optických komunikačných systémov sa využíva optický variant časového multiplexu nazývaný multiplex s opticko-časovým delením OTDM (Optical Time Division Multiplex). Zjednodušená bloková schéma multiplexu OTDM je zobrazená na Obr. 1 [1, 2, 3].



Obr. 1 Multiplex s opticko-časovým delením.

Pri optickom multiplexe s časovým delením OTDM vchádzajú vysielané optické signály do zariadenia nazývaného multiplexor. Ten každý vstupný optický signál rozdelí na krátke impulzy, ktorým priradí časové úseky. Priradené časové úseky určuje perióda, počas ktorej sa môžu dané optické signály vysielat'. Je dôležité zabezpečiť, aby sa jednotlivé časové úseky navzájom neprekrývali a neovplyvňovali. Preto sa medzi jednotlivé úseky pomocou oneskorenia vkladá ochranné časové pásmo. Podľa toho, či je časový úsek priradzovaný pevne alebo dynamicky rozoznávame dve metódy TDM [7][8]:

- Synchronne TDM (Synchronous TDM). Časové úseky sú priradzované pevne. Rovnakej vysielacej stanici je vždy pridelené rovnaké časové okno. Nevýhodou je, že v prípade, keď jedna vysielacia stanica nevysiela, ostane jedno časové okno nevyužitú.
- Asynchronne TDM (Asynchronous TDM). Časové úseky sú priradzované dynamicky. Rovnakej vysielacej stanici nemusí byť vždy pridelené rovnaké časové okno. Úseky sú priradzované vysielacím staniciam podľa toho, ktoré časové okno je v danom momente voľné. Výhodou je efektívnejšie využívanie zdieľaného komunikačného kanála.

Optické signály sú následne prenášané zdieľaným komunikačným kanálom, resp. prenosovým médiom v podobe optického vlákna na jednej vlnovej dĺžke. Na prijímacej strane vchádzajú optické signály do zariadenia nazývaného demultiplexor. Ten jednotlivé optické signály rozdeľuje, zostavuje ich do pôvodnej formy a posieľa do konečných staníc podľa cieľovej destinácie. Vďaka veľkej šírke pásma a prenosovej rýchlosti zdieľaného komunikačného kanála sa vysielané optické signály malej veľkosti javia, ako keby boli prenášané paralelne. Zdieľaný komunikačný kanál vieme opísať nasledujúcou rovnicou [2, 3]:

$$R = C * S * n, \quad (1)$$

pričom rýchlosť komunikačného kanála R je daná počtom multiplexovaných kanálov (vysielacích staníc) C , počtom vzoriek prenesených za jednu sekundu S a počtom bitov n na jednu prenesenú vzorku. V súčasnosti vieme pomocou optických sietí na báze TDM prenášať dáta v rýchlostiach na úrovni 10 Gbit/s a viac [4, 5, 6].

A. Gigabitová pasívna optická sieť GPON

S GPON štandardom podľa ITU-T G.984 prišlo v r. 2003 združenie poskytovateľov služieb v prístupových sieťach FSAN (Full Service Access Network), v ktorom definovalo úplne nový typ PON siete v podobe gigabitovej pasívnej optickej siete. Podpora prenosu dát metódou ATM (štandardy APON a BPON) zostala aj naďalej podporovaná. Rozširovaním protokolu Ethernet (EPON) sa ale zaviedol aj nový spôsob prenosu a to pomocou protokolu založenom na „zapuzdrovaní dát“ tzv. GPON rámcov GEM (GPON Encapsulation Method). Metóda GEM tak poskytuje koncovým užívateľom okrem vyššej prenosovej rýchlosti aj lepšiu kvalitu služieb QoS (Quality of Services) v dátových prenosoch, ktoré sú citlivé na oneskorenia (napr. živé vysielanie, hranie online hier alebo telefonovanie).

Tabuľka 1
Porovnanie základných parametrov PON sietí prvej generácie

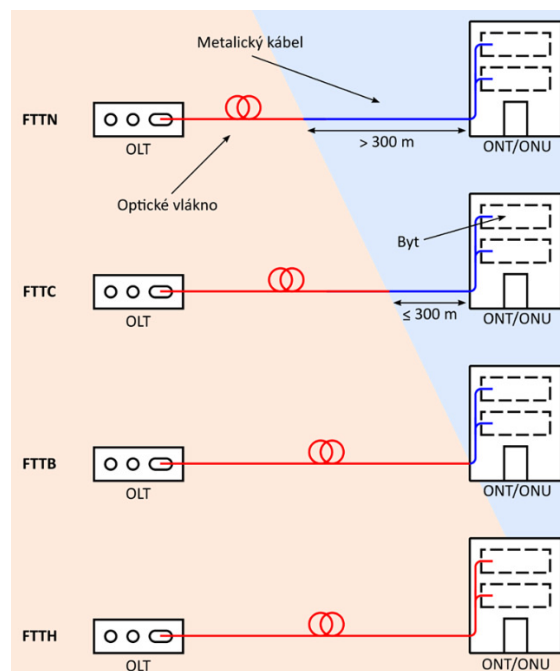
	APON	BPON	EPON	GPON
Rok	1998	2001	2004	2003
Štandard	ITU-T G.983	ITU-T G.983.1	IEEE 802.3ah	ITU-T G.984
Protokol	ATM	ATM	Ethernet	ATM, GEM
Počet koncových účastníkov	32	32	32	128
Vlnové dĺžky vo vzostupnom smere [nm]	1260 – 1360	1260 – 1360	1260 – 1360	1260 – 1360
Vlnové dĺžky v zostupnom smere [nm]	1480 – 1580	1480 – 1580	1480 – 1500	1480 – 1550
Prenosová rýchlosť vo vzostupnom smere [Mbit/s]	155,52	622,08	1244,16	1244,16
Prenosová rýchlosť v zostupnom smere [Mbit/s]	622,08	1244,16	1244,16	2488,32

Maximálna prenosová rýchlosť GPON sietí v symetrickom smere predstavuje 1244,16 Mbit/s. Aj napriek tomu, že prenosová rýchlosť 1 Gbit/s bola prekonaná už štandardom BPON, príchodom najnovšieho GPON štandardu ITU-T G.984.7 bola v sieťach GPON táto hranica

prekonaná viac ako dvojnásobne. Najnovšie GPON siete dosahujú pri asymetrickom prenose prenosové rýchlosti 155,52 Mbit/s, 622,08 Mbit/s alebo 1244,16 Mbit/s vo vzostupnom smere a 1244,16 Mbit/s alebo 2488,32 Mbit/s v zostupnom smere. Prenosové rýchlosti v GPON sieťach sú teda vysoko flexibilné a podľa potreby je jednoduché ich prispôbiť. Dnešné GPON siete využívajú vo vzostupnom smere vlnovú dĺžku 1310 nm a v zostupnom smere vlnovú dĺžku 1550 nm. Na správu siete sa využíva asi 7% celkovej prenosovej kapacity. Okrem vylepšených prenosových rýchlostí a QoS je novinkou v GPON sieťach aj možnosť distribúcie prepínanej IPTV. Maximálna prenosová vzdialenosť štandardných GPON sietí ostala rovnaká ako pri sieťach APON a BPON, teda približne 20 km od OLT k ONT/ONU. Výnimkou sú GPON siete s predĺženým dosahom nazývané RE GPON (Reach Extended GPON) alebo LR GPON (Long Reach GPON), v ktorých sa maximálna prenosová vzdialenosť zvyšuje až na 60 km. Rozbočovací pomer v GPON sieťach bol rovnako zvýšený, a to z pôvodných 32 na 128 koncových zariadení. Siete GPON dnes vďaka vysokým prenosovým rýchlostiam, pomerne nízkym nákladom a jednoduchšej údržbe patria k najrozšírenejším PON sieťam. Základné špecifikácie PON sietí prvej generácie sú zhrnuté v Tabuľke 1.

B. Prístupová metóda PON typu FTTH

Optické prístupové siete PON vieme okrem využitého rýchlostného štandardu ďalej rozdeľovať aj na základe spôsobu pripojenia koncových užívateľov a vzdialenosti ukončenia optického vlákna od koncového užívateľa. Spoločné skupinové označenie prístupových metód v PON sieťach sa označuje ako FTTx (Fiber To The „x“) siete. Za písmeno „x“ v názve sa dosadzuje identifikátor určujúci vzdialenosť miesta ukončenia optického vlákna od koncového užívateľa. Pomocou prístupových metód FTTx a spojení P2MP je možné efektívne distribuovať signál prístupovou sieťou PON a poskytovať tak vysokú kvalitu triple-play služieb. Optické vlákno môže byť v prístupových optických sieťach ukončené priamo v dome, resp. byte koncového užívateľa, v rozvádzači umiestnenom v bytovom dome alebo v samostatnom optickom rozvádzači umiestnenom vonku. Existuje mnoho rozličných kombinácií a špecificky upravených prístupových metód. Grafické znázornenie vybraných prístupových metód FTTx je na Obr. 2.



Obr. 2 Prístupové metódy FTTx.

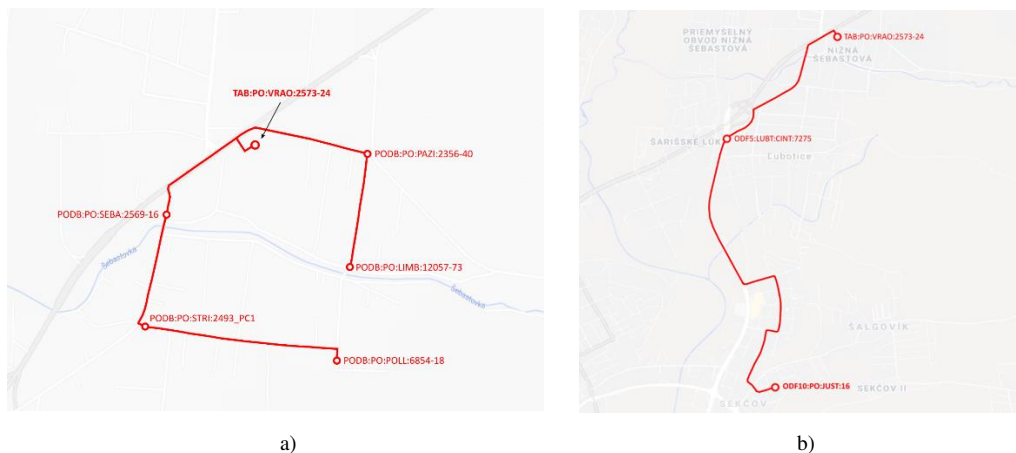
Metóda optické vlákno do domu FTTH (Fiber To The Home) v súčasnosti predstavuje aj napriek náročnosti výstavby najdokonalejší a zároveň najrozšírenejší spôsob pripájania koncových účastníkov do optickej siete. Optické vlákno je pri tejto metóde vedené z telekomunikačnej ústredne do optického rozvádzača. V ňom je rozdelené pomocou rozbočovača na viacero optických vlákien, ktoré sú následne vedené priamo do domov či bytov koncových užívateľov, kde sú ukončované v zariadeniach ONT/ONU bez využívania metalických káblov. Prenosové rýchlosti sú poskytované v asymetrickom režime, keďže väčšina koncových účastníkov si žiada rýchlejší download (sťahovanie) ako upload (nahrávanie) dát. Metóda FTTH

je teda čisto optická a odstraňuje tzv. efekt hrdla fľaše (bottleneck), ktorý pri opticko-metalických prístupových metódach znižuje maximálne prenosové rýchlosti pri konvertovaní optického signálu na metalický. Nahradenie metalických káblov optickými vláknami tiež umožňuje plné využitie rýchlostných štandardov EPON, GPON či dokonca NG-PON2.

III. NÁVRH A REALIZÁCIA FTTH SIETE PREŠOV – NIŽNÁ ŠEBASTOVÁ

A. Návrh FTTH siete

Pre primárnu časť siete boli navrhnuté 4 primárne uzlové body vo forme PO DB (Pasívny Optický Distribučný Bod). Každé označenie PO DB, ale aj iných uzlových bodov, obsahuje skratku obce, ulice, súpisné a orientačné číslo domu, resp. parcely, pri ktorej sa nachádza. Primárne uzlové body PO DB boli umiestnené na uliciach Limbová (PO DB:PO:LIMB:12057-53), Pažica (PO DB:PO:PAZI:2356-40), Strážnická (PO DB:PO:STRI:2493_PC1) a Poľná (PO DB:PO:POLL:6854-18). Počas prác sa však zistilo, že z dôvodu napojenia sekundárnej časti siete na Južnej ulici je nutné umiestniť aj piaty uzlový bod na Šebastovskej ulici (PO DB:PO:SEBA:2569-16). Prívod optického signálu do PO DB je riešený pomocou transportnej optickej siete začínajúcej na Justičnej ulici v mestskej časti Prešov – Sekčov zo zdroja umiestneného v rozvážači ODF (Optical Distribution Frame) (Obr. 3a).



Obr. 3 (a) FTTH sieť Prešov – Nižná Šebastová – trasa primárnej časti siete, (b) Transportná sieť Prešov – Nižná Šebastová – trasa siete.

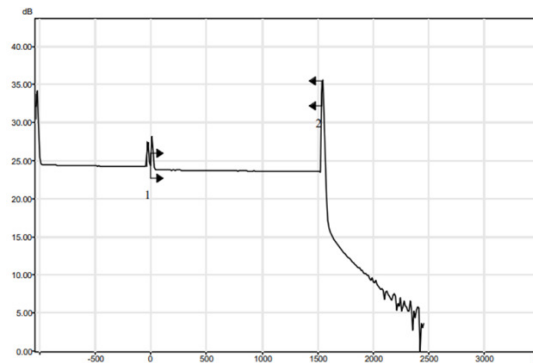
Transportná sieť končí v bode TAB (Transportný Agregáčny Bod) na Vranovskej ulici pri obchodnom dome – TAB:PO:VRAO:2573-24 (Obr. 3b). Bod TAB slúži ako brána medzi transportnou a prístupovou sieťou. Na Obr. 3a je červenou farbou znázornená trasa primárnej časti siete spolu s umiestnením bodu TAB a všetkých PO DB. Pre sekundárnu časť siete bolo navrhnutých 157 ZB (Zákaznícky Bod), 5 SPODB (Stĺpový Pasívny Optický Distribučný Bod) a z dôvodu veľkého množstva bytových domov až 19 ChPODB (Chodbový Pasívny Optický Distribučný Bod). Každý z týchto bodov poskytuje pripojenie 2 až 12 koncovým účastníkom. Sekundárna časť siete pokrýva takmer všetky byty, kancelárske priestory, parcely a rodinné domy mestskej časti. Výnimkou sú len pozemky, ktorých majitelia nesúhlasili so zriadením optickej prípojky. Finálna podoba FTTH siete Prešov – Nižná Šebastová je zobrazená na Obr. 4. Červenou farbou je zvýraznená trasa primárnej časti, modrou farbou trasa sekundárnej časti siete.



Obr. 4 FTTH sieť Prešov – Nižná Šebastová – trasa primárnej a sekundárnej časti siete.

B. OTDR merania FTTH siete Prešov – Nižná Šebastová

Jednou z podmienok pre úspešné ukončenie prác na optickej sieti je odovzdanie meracích protokolov, ktorých výsledky sú v súlade s požiadavkami kladenými na danú sieť operátorom. Po úspešnom dokončení prác na každej jednej vetve sa pomocou meracích prístrojov opticko-časovej reflektometrie OTDR (Optical Time Domain Reflectometry) získali krivky spätného rozptylu. Na Obr. 5 je ako príklad zobrazená krivka spätného rozptylu pre najdlhší úsek siete – PO DB:PO:POL L:6854-18. Dôvodom náhlych skokov pri bodoch 1 a 2 sú odrazy od začiatku a konca optickej trasy.



Obr. 5 FTTH sieť Prešov – Nižná Šebastová – PO DB:PO:POL L:6854-18 – krivka spätného rozptylu – vlnová dĺžka 1550 nm.

Všetky uzlové body PO DB v FTTH sieti Prešov – Nižná Šebastová sú napájané pomocou vláknových zväzkov obsahujúcich 6 až 12 jednovidových vlákien SMF (Single Mode Fiber). Merania sa vždy vykonávajú pre dve vlnové dĺžky 1310 nm (Tabuľka 2) a 1550 nm.

Tabuľka 2

FTTH sieť Prešov – Nižná Šebastová – výsledky OTDR merania na reálnej sieti – vlnová dĺžka 1310 nm

Miesto merania	LIMB: 12057-73	PAZI: 2356-40	SEBA: 2569-16	STRI: 2493_PC1	POLL: 6854-18
Priemer a typ vlákna	9 μm SM	9 μm SM	9 μm SM	9 μm SM	9 μm SM
Dĺžka vlákna [km]	0,7939	0,4487	0,3533	0,7590	1,5332
Priemerný útlm [dB.km⁻¹]	0,346	0,304	0,379	0,343	0,347
Celkový útlm [dB]	0,275	0,136	0,134	0,260	0,532
ORL [dB]	36,35	37,44	34,66	33,42	31,60
Doba merania [s]	15	15	15	15	15
Šírka impulzu [ns]	30	30	100	100	100
Rozsah merania [km]	2,5	2,5	2,5	2,5	3,5

Z Tabuľky 2 vyplýva, že pri vlnovej dĺžke 1310 nm bol najvyšší celkový útlm, s hodnotou 0,532 dB, nameraný na úseku PO DB:PO:POL L:6854-18. Táto hodnota zároveň zodpovedá skutočnosti, že úsek má najväčšiu dĺžku optického vlákna, a to 1,5332 km. Najnižšia hodnota celkového útlmu pri vlnovej dĺžke 1310 nm bola nameraná na úseku PO DB:PO:SEBA:2569-16, ktorý je zároveň najkratším úsekom primárnej časti testovanej prístupovej siete s dĺžkou 0,379 km. Ďalším zaujímavým parametrom je priemerný útlm daných úsekov. Jeho hodnota závisí od dĺžky optického vlákna a vlnovej dĺžky. Priemerný útlm sa vo všetkých úsekoch pri vlnovej dĺžke 1310 nm pohybuje v rozmedzí od 0,304 do 0,379 dB.km⁻¹, pričom najvyšší bol nameraný na úseku PO DB:PO:SEBA:2569-16. Najnižší priemerný útlm bol nameraný na úseku PO DB:PO:PAZI:2356-40 aj napriek tomu, že sa nejedná o najkratší úsek siete. Parameter odrazu útlmu ORL (Optical Return Loss) je ďalším parametrom, pomocou ktorého je možné vyhodnotiť kvalitu realizovanej siete. Pri OTDR meraniach udáva hodnotu optického výkonu, ktorý sa vráti späť na zdroj kvôli odrazom spôsobeným na trase.

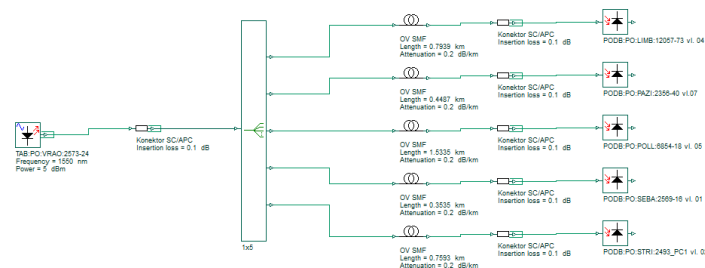
Optické siete vysielajúce na vlnovej dĺžke 1310 nm štandardne dosahujú oproti vlnovej dĺžke 1550 nm vyššie celkové útlmy. Na krátkych úsekoch, akými boli časti primárnej siete FTTH Prešov – Nižná Šebastová sa táto vlastnosť neprejavila tak významne, ako by sa prejavila na väčších vzdialenostiach. Štatistické porovnanie priemerných útlmov pre spomínané vlnové dĺžky na optickom vlákne SMF je zobrazené v Tabuľke 3.

Tabuľka 3
Štatistické porovnanie útlmov optického vlákna SMF – vlnové dĺžky 1310 nm a 1550 nm (hodnoty v dB)

	500 m	2 km	10 km	40 km	80 km
1310 nm	0,21	0,84	4,2	16,8	33,6
1550 nm	0,14	0,56	2,8	11,2	22,4

C. Simulácia FTTH siete Prešov – Nižná Šebastová v prostredí programu OptiSystem

Na základe hodnôt a parametrov získaných z OTDR meraní na reálnej optickej linke a projektovej dokumentácie siete bol v prostredí simulačného programu OptiSystem navrhnutý experimentálny model FTTH siete Prešov – Nižná Šebastová. Všetky vetvy siete sú v skutočnosti napájané samostatným optickým vláknom z jedného zdrojového bodu TAB. Na žiadnej z liniek sa v skutočnosti rozbočovač nenachádza, no pre zjednodušenie simulačného modelu bol do schémy jeden pridaný s nastaveným nulovým vloženým útlmom. Finálna podoba simulačného modelu je znázornená na Obr. 6.



Obr. 6 FTTH sieť Prešov – Nižná Šebastová – simulačný model siete v programe OptiSystem.

D. Cenová bilancia FTTH siete Prešov – Nižná Šebastová

Cenová bilancia vychádza z projektovej dokumentácie, ktorá obsahuje databázu materiálov a montážnych či výkopových prác. Výsledná cenová bilancia je orientačná a nemusí predstavovať skutočné náklady na vybudovanie prístupovej siete, pretože sa v nej nenachádzajú niektoré položky, ktoré sa pri realizácii siete dodatočne použili. Výsledná suma za materiály predstavuje hodnotu cca 58 000 €. Suma je vypočítaná na základe dostupných predajných cien a počtu merných jednotiek daných materiálov. Cena za montážne a výkopové práce je oproti cene za materiály z pravidla podstatne vyššia. Vzhľadom na skutočnosť, že značná časť trasy siete bola realizovaná v chodníkoch môže predstavovať cena za realizačné práce sumu cca 250 000 €. Preto odhadujeme, že finálna cenová bilancia FTTH siete Prešov – Nižná Šebastová je približne 310 000 €. Finálna suma však neobsahuje náklady na projektové práce.

IV. ZÁVER

Vzhľadom na výsledky, ktoré dosahuje GPON sieť z pohľadu prenosových rýchlostí je otázkou času, kedy táto sieť plne nahradí metalické siete ADSL a VDSL, rovnako ako aj rádiové, či dokonca satelitné technológie zaoberajúce sa prenosom telekomunikačného signálu. Vďaka tomu sú poskytovatelia služieb nútení na tento súčasný trend reagovať a svoje služby pripravovať na nástup optických sietí. Tie okrem zvýšenej rýchlosti výrazne navýšia odozvu siete, čo je výhodné pre domácich účastníkov siete sledujúcich televíziu alebo pre skvalitnenie technológie IoT (Internet of Things) v mestách a rozľahlých regiónoch.

POĎAKOVANIE

Táto práca bola podporená výskumným projektom FEI-2022-84 "Techniky spracovania dát vo vysokorýchlostných prenosových systémoch".

LITERATÚRA

- [1] V. Sharma, et al., "Passive Optical Network: A New Approach In Optical Network", In: *2020 International Conference on Advances in Computing, Communication & Materials (ICACCM)*, IEEE 2020, pp. 295-300. doi:10.1109/ICACCM50413.2020.9213059
- [2] ITU-T G.984.1. Gigabit-capable passive optical networks (GPON): General characteristics. ITU-T. March 2008.
- [3] D&O Committee. 2017. FTTH Handbook. Edition 8. FTTH Council Europe. 204 pp. Revision 2018-03-13.
- [4] ITU-T G.987.3. 10-Gigabit-capable passive optical networks (XG-PON): Transmission convergence (TC) layer specification. ITU-T. January 2014.
- [5] ITU-T G.989. 40-Gigabit-capable passive optical networks (NG-PON2): Definitions, abbreviations and acronyms. ITU-T. October 2015.
- [6] Calix. Next-Generation PON: Eliminating Physical constraints from the access network, 2017, https://www.calix.com/content/dam/calix/marketing-documents/public/Calix_Technical_Brief_NG-PON2.pdf

SAT kódovanie kardinalitných ohraničení pomocou P-H metódy

¹Lubomír RUSNÁK, ²Marián MACH

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

¹lubomir.rusnak@student.tuke.sk, ²marian.mach@tuke.sk

Abstrakt — Článok sa zameriava na problematiku vytvárania logických modelov úloh prostriedkami výrokovkej logiky keď hľadanie riešenia úlohy je ekvivalentné nájdeniu podmienok splniteľnosti vytvoreného modelu. Pozornosť je zameraná na kardinalitné ohraničenia obmedzujúce počet vybraných alternatív z dostupnej množiny alternatív. Článok sa zaoberá modelovaním takýchto ohraničení v tvare CNF vytváranom nad logickými symbolmi pomocou negácie, konjunkcie a disjunkcie kódovaním kardinalitných ohraničení, pričom sa hlbšie venuje P-H (Pigeon-Hole) metóde kódovania. V rámci experimentovania bolo realizovaných niekoľko experimentov s viacerými variantmi tohto kódovania a ich porovnaním voči známemu binomickému kódovaniu v rámci zvoleného testovacieho problému.

Kľúčové slová — binomické kódovanie, kardinalitné ohraničenia, logický model, pigeon-hole kódovanie, výroková logika

I. ÚVOD

Pri hľadaní exaktného (nie aproximatívneho) riešenia nejakého problému je vhodné zvážiť aj použitie prístupu založeného na logickom modelovaní. Populárnym v tejto oblasti je použitie výrokovkej logiky, ktorá ponúka logické symboly a zbierku unárnych a binárnych logických operátorov (negácia, konjunkcia, disjunkcia, ...) ako základné stavebné kamene. Ich vhodnou voľbou je možné zostaviť logické výroky a na nich vybudovať logický model úlohy, ktorý je následne testovaný na validnosť alebo splniteľnosť [1]. V prípade splniteľnosti hodnoty symbolov, ktoré vedú na splniteľný model, vlastne priamo kódujú hľadané riešenie.

Pri vytváraní logického modelu sa typicky reprezentujú rozličné ohraničenia, ktoré musí model spĺňať. Častým prípadom takýchto ohraničení sú kardinalitné ohraničenia [2]. Tieto ohraničenia modelujú situáciu, keď existuje viac alternatív, z ktorých možno validne vybrať iba určitý počet – výber iného počtu je považovaný za nevalidný. Najčastejší tvar kardinalitných ohraničení je voľba

- maximálne b alternatív,
- minimálne b alternatív,
- presne b alternatív,

z celkového počtu n alternatív ($0 < b < n$). Tieto alternatívy je možné nahradiť všeobecnejším tvarom výberu “minimálne l_1 a súčasne maximálne l_2 alternatív z počtu n alternatív” ktorý vhodnou voľbou hodnôt l_1 a l_2 umožňuje reprezentovať všetky tri vyššie uvedené formy ohraničenia. Na druhej strane medzi týmito tromi formami ohraničení existujú vzájomné vzťahy, umožňujúce ich vzájomnú zámenu (napríklad “minimálne b z n alternatív” je to isté ako “maximálne $n-b$ z n negovaných alternatív”).

Problémom je, že SAT solver (softvér používaný na dokazovanie splniteľnosti logických modelov) takto definovaným ohraničeniam nerozumie. Preto je potrebné takéto vysokoúrovňové vyjadrenie kardinalitných ohraničení zakódovať do (typicky) tvaru logického výrazu reprezentovaného vo forme CNF (konjunktívnej normálnej formy), kde jednotlivé alternatívy sú reprezentované logickými symbolmi a samotné ohraničenie medzi nimi výrazom zostaveným pomocou troch operátorov – negácie, disjunkcie a konjunkcie [3].

Tento článok sa zameriava na jednu metódu kódovania kardinalitných ohraničení. Jeho štruktúra je nasledovná – v druhej kapitole sa venuje metóde kódovania v porovnaní s metódou, ktorá je v danej oblasti považovaná za základnú a v tretej kapitole je uvedená séria experimentov s danými metódami na úlohe hľadania riešenia nesplniteľného problému.

II. METÓDY KÓDOVANIA KARDINALITNÝCH OHRANIČENÍ

A. Binomické kódovanie

Táto metóda je najjednoduchšou kódovacou metódou a preto je často (ako aj v našom prípade) používaná ako porovnávacia metóda. Je zameraná na ohraničenie tvaru

$$\leq_b(X_1, X_2, \dots, X_n) \quad (1)$$

teda počet alternatív, vybraný z celkového počtu n dostupných alternatív X_1, X_2 až X_n je menší alebo rovný prahovej hodnote b .

Keďže naraz môže byť maximálne b symbolov splnených, tak pre žiadnu podmnožinu $b+1$ symbolov nemôže platiť, že by boli všetky splnené – aspoň jeden z tejto $(b+1)$ -tice musí byť nespĺnený. V prípade $b=1$ je výsledok kódovania

$$\bigwedge_{i=1}^{n-1} \bigwedge_{j=i+1}^n (\neg X_i \vee \neg X_j) \quad (2)$$

kde výraz v zátvorkách reprezentuje klauzulu, ktorá hovorí že aspoň jeden z uvedených symbolov nesmie byť splnený.

Binomické kódovanie je jediným kódovaním, ktoré používa iba tie symboly, ktoré priamo reprezentujú dostupné alternatívy. Jeho nevýhodou je však často veľký počet generovaných klauzúl, preto sú vytvárané ďalšie metódy kódovania, ktoré sa snažia zmenšiť počet generovaných klauzúl aj za cenu zavádzania dodatočných pomocných symbolov.

B. P-H (Pigeon-Hole) kódovanie

Na rozdiel od predchádzajúcej metódy, táto metóda [4] je zameraná na kódovanie ohraničenia tvaru

$$\geq_b(X_1, X_2, \dots, X_n) \quad (3)$$

teda počet alternatív, vybraný z celkového počtu n dostupných alternatív X_1, X_2 až X_n je väčší alebo rovný prahovej hodnote b .

Používa sadu pomocných symbolov. Pre vysvetlenie idey, na ktorej je metóda založená, uvažujme usporiadanie pomocných symbolov v tvare matice

$$\begin{bmatrix} p_{1,1} & p_{1,2} & p_{1,3} & p_{1,4} & p_{1,5} \\ p_{2,1} & p_{2,2} & p_{2,3} & p_{2,4} & p_{2,5} \\ p_{3,1} & p_{3,2} & p_{3,3} & p_{3,4} & p_{3,5} \end{bmatrix} \quad (4)$$

kde každý stĺpec zodpovedá jednému zo symbolov X_i (teda počet stĺpcov je n) a počet riadkov je daný hodnotou prahu b . Keďže pre splnenie ohraničenia stačí vybrať b symbolov, tak je potrebné realizovať b výberov – a teda každému výberu zodpovedá jeden riadok pomocných symbolov. Symbol $p_{i,j}$ sa viaže na vybratie j -teho symbolu v i -tom výbere. Je zrejmé, že matica uvedená v (4) zodpovedá ohraničeniu $\geq_3(X_1, X_2, X_3, X_4, X_5)$, teda $n=5$ a $b=3$.

Pre správne kódovanie kardinalitného ohraničenia je potrebné splniť tri podmienky, ktoré vlastne dané ohraničenie vyjadria v tvare logických výrokov nad použitými symbolmi.

1. Ak nejaký pomocný symbol v niektorom stĺpci je splnený, tak splnená je aj alternatíva viažúca sa s daným stĺpcom.
2. Aby každý výber znamenal vybratie aspoň jednej alternatívy, tak v každom riadku musí byť splnený aspoň jeden pomocný symbol.
3. Aby bolo zabezpečené vybratie aspoň b alternatív, tak rôzne výbery musia zaručovať vybratie rôznych alternatív – teda dva výbery nemôžu vybrať tú istú alternatívu. Teda v každom stĺpci môže byť splnený najviac jeden pomocný symbol.

Tieto tri podmienky pre príklad uvedený v (4) je možné kódovať pomocou nasledovných logických výrokov:

$$\bigwedge_{i=1}^3 \bigwedge_{j=1}^5 (\neg p_{i,j} \vee X_j) \quad (5)$$

$$\bigwedge_{i=1}^3 (p_{i,1} \vee p_{i,2} \vee p_{i,3} \vee p_{i,4} \vee p_{i,5}) \quad (6)$$

$$\bigwedge_{j=1}^5 \bigwedge_{i=1}^2 \bigwedge_{k=i+1}^3 (\neg p_{i,j} \vee \neg p_{k,j}) \quad (7)$$

Takto je definovaný prvý variant P-H metódy kódovania, ktorý budeme v ďalšom značiť PH1. Druhý variant PH2 vychádza z PH1 s jedinou modifikáciou – zníženie počtu pomocných symbolov tak, aby symboly boli usporiadané podľa

$$\begin{bmatrix} & p_{1,3} & p_{1,4} & p_{1,5} \\ & p_{2,2} & p_{2,3} & p_{2,4} \\ p_{3,1} & p_{3,2} & p_{3,3} & \end{bmatrix} \quad (8)$$

kde musia platiť a byť kódované presne tie isté tri podmienky ako pre PH1.

Tretí variant PH3 používa tie isté pomocné symboly ako PH2 tiež v usporiadaní podľa (8), avšak podstata je v zmene kódovania tretej podmienky aby rôzne výbery (riadky) vybrali rôzne alternatívy. Namiesto kódovania

$$(\neg P_{2,2} \vee \neg P_{3,2}) \wedge (\neg P_{1,3} \vee \neg P_{2,3}) \wedge (\neg P_{1,3} \vee \neg P_{3,3}) \wedge (\neg P_{2,3} \vee \neg P_{3,3}) \wedge (\neg P_{1,4} \vee \neg P_{2,4}) \quad (9)$$

teraz použije kódovanie tretej podmienky pomocou

$$(P_{2,2} \vee \neg P_{1,3}) \wedge (P_{2,2} \vee P_{2,3} \vee \neg P_{1,4}) \wedge (P_{3,1} \vee \neg P_{2,2}) \wedge (P_{3,1} \vee P_{3,2} \vee \neg P_{2,3}) \quad (10)$$

kde kódovanie vzájomného vylučovania sa dvojíc symbolov bolo nahradené implikačnými vzťahmi ktoré lepšie zodpovedajú vnútornému spôsobu inferencie, ktorý je realizovaný v SAT solveroch.

Štvrtý variant PH4 kóduje podmienky rovnakým spôsobom ako PH3, avšak znižuje počet potrebných pomocných symbolov, pretože maticu (8) modifikuje do tvaru

$$\begin{bmatrix} & p_{1,3} & p_{1,4} & x_5 \\ & p_{2,2} & p_{2,3} & x_4 \\ p_{3,1} & p_{3,2} & x_3 & \end{bmatrix} \quad (11)$$

pričom niektoré pomocné symboly nahradí priamo symbolmi alternatív, ktoré sú k dispozícii. Súčasne sa zníži aj počet generovaných klauzúl, pretože napríklad prvú z troch kódovaných podmienok nie je potrebné zapisovať pre posledný stĺpec, pretože $\neg X_5 \vee X_5$ je tautológiou. Podobné tautológie nie je potrebné generovať ani v niekoľkých ďalších stĺpcoch.

Piaty variant PH5 vychádza z PH4 a ešte agresívnejším spôsobom nahrádza pomocné symboly tými symbolmi, ktoré reprezentujú dostupné alternatívy. Maticu (11) modifikuje na tvar

$$\begin{bmatrix} & p_{1,3} & p_{1,4} & x_5 \\ & p_{2,2} & p_{2,3} & x_4 \\ x_1 & x_2 & x_3 & \end{bmatrix} \quad (12)$$

pričom opäť dochádza k ďalšej úspore počtu generovaných klauzúl. Existujú aj ďalšie optimalizácie, ktorým sme sa však už nevenovali [5].

III. VYBRANÉ EXPERIMENTY

A. Testovací problém

Pre experimentovanie s rôznymi metódami kódovania kardinalitných ohraničení bol zvolený problém ukladania predmetov do krabíc. Pri tomto probléme sú zadané dve množiny objektov – p predmetov nejakého typu a h krabíc. Krabice majú rovnakú kapacitu c – do každej krabice je možné vložiť maximálne c predmetov.

Úlohou je uložiť všetky predmety do krabíc tak, aby boli dodržané nasledovné dve podmienky:

- Každý predmet má byť vložený do práve jednej krabice – žiadny predmet nemôže ostať mimo a taktiež žiadny predmet nemôže byť vložený súčasne do viacerých krabíc.
- Je potrebné rešpektovať maximálnu kapacitu krabíc – do každej krabice môže byť vložených maximálne c predmetov.

Je zrejmé, že v prípade $p \leq c \cdot h$ úloha má riešenie (vlastne väčší počet riešení vzhľadom na symetriu medzi krabicami a symetriu medzi predmetmi), zatiaľ čo v prípade $p > c \cdot h$ žiadne riešenie neexistuje.

Aby bolo možné porovnať nielen výsledné logické modely generované rôznymi metódami kódovania ale aj ich vplyv na riešenie SAT solverom, bol zvolený nesplniteľný variant úlohy s neexistujúcim riešením. V tomto prípade totiž solver musí najprv prehľadať celý priestor možných riešení a až po jeho úplnom prehľadaní je schopný konštatovať nesplniteľnosť logického modelu. Pre tento variant je možné upraviť prvú podmienku na jednoduchší tvar – každý predmet musí byť vložený do minimálne jednej krabice. Táto úprava zjednoduší model a zároveň nezmení nesplniteľný model na splniteľný.

Model riešeného príkladu je teda tvorený dvomi kardinalitnými ohraničeniami. Ak symbol $X_{i,j}$ bude reprezentovať vloženie i -teho predmetu do j -tej krabice (celkovo je potrebných $p \cdot h$ takýchto symbolov), tak tieto ohraničenia budú mať tieto tvary

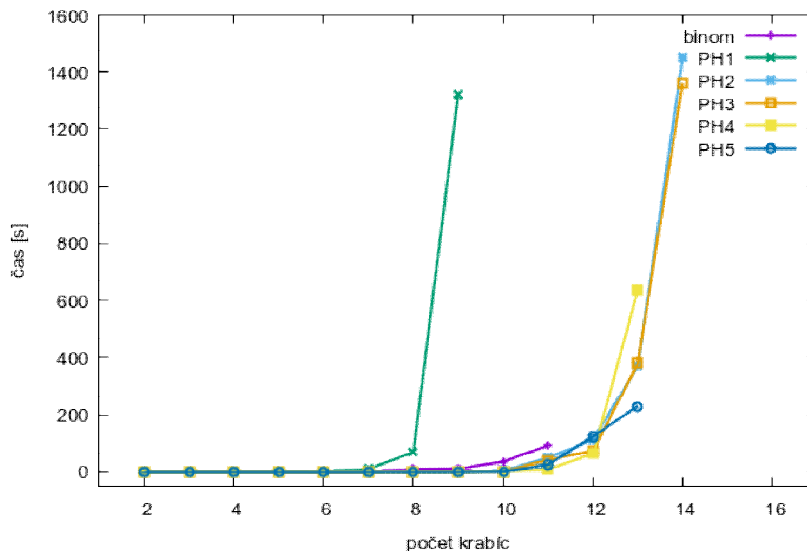
$$\forall_{i=1}^p \geq_1 (X_{i,1}, X_{i,2}, \dots, X_{i,h}) \quad (13)$$

$$\forall_{j=1}^h \leq_c (X_{1,j}, X_{2,j}, \dots, X_{p,j}) \quad (14)$$

B. Časová zložitosť generovaných modelov

V tomto experimente boli použité všetky metódy kódovania spomenuté v predchádzajúcej kapitole [6]. Boli generované výsledné logické modely a tie vložené do SAT solvera, ktorý vo všetkých prípadoch dokázal nesplniteľnosť týchto modelov čím overil ich správnosť.

Modely boli generované pre rôzny počet krabíc – najmenšia hodnota bola 2 a postupne sa inkrementálne zvyšovala až dovedy, kým solver bol neschopný spracovať model v rámci doby 30 minút. Výsledky sú znázornené na obr. 1.

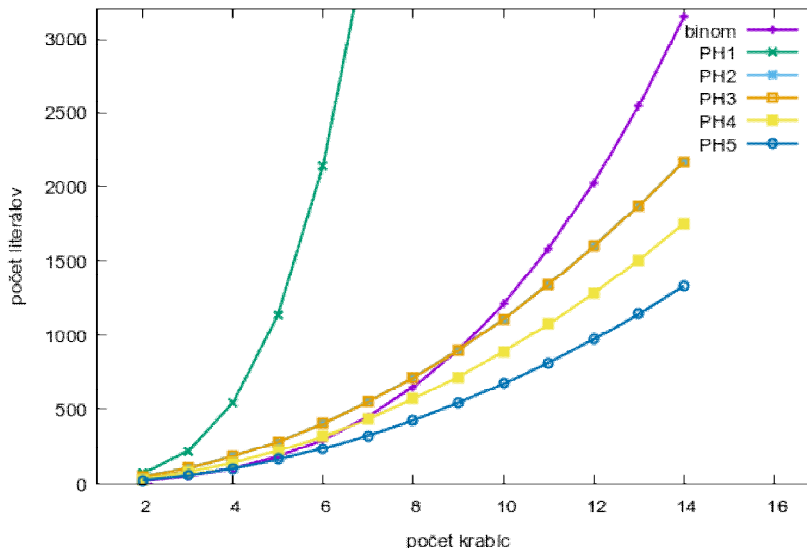


Obr. 1 Čas potrebný pre dokázanie nesplniteľnosti

Binomické kódovanie dokázalo vo vymedzenom čase zvládnuť model s 11 krabicami, časový nárast pre spracovanie modelu s 12 krabicami už presiahol vymedzený čas. Je zrejmé, že metóda PH1 dosiahla najhoršie výsledky čo neprekvapuje, pretože jednoducho pridávala pomocné symboly bez akejkoľvek optimalizácie. Optimalizované podoby PH metódy si viedli lepšie – vo vymedzenom čase dokázali spracovať modely väčšie o dve alebo tri krabice oproti výkonu binomického kódovania. Pritom o niečo lepšie si viedli metódy s väčším počtom pomocných symbolov (PH2 a PH3) oproti metódam s asymetrickou náhradou niektorých pomocných symbolov priamo symbolmi kódujúcimi jednotlivé alternatívy (PH4 a PH5). Prechod z kódovania vylučujúcich sa dvojíc (PH2) na kódovanie implikačných vzťahov (PH3) sa ukázal ako prospešný.

C. Veľkosť generovaných modelov

Jednotlivé kódovacie metódy sú schopné generovať logické modely rôznych veľkostí. Jednou z mier pre popis veľkosti modelov je počet literálov (pozitívnych alebo negovaných symbolov) použitých v zostave modelu. Výsledky sú graficky znázornené na obr. 2.



Obr. 2 Veľkosť logických modelov podľa počtu použitých literálov

Najväčšie modely generuje PH1 s veľmi rýchlym nárastom počtu literálov (až so 47670 literálmi pre model so 14 krabicami). Zrejme tento veľký počet literálov sa podieľal na vyššej časovej zložitosti pri spracovaní takto kódovaných modelov. Ostatné PH metódy (PH2 a PH3 sa navzájom prekrývajú) podľa očakávania produkujú menšie modely než binomické kódovanie. Zároveň platí že čím menší počet doplnených pomocných symbolov, tým menší aj výsledný logický model. Veľkosť generovaných modelov dobre korešponduje s celkovým počtom pridávaných pomocných symbolov, ktorý je v tab. 1.

Tabuľka 1
Celkový počet použitých pomocných symbolov.

Počet krabic	PH1	PH2	PH3	PH4	PH5
2	18	14	14	7	2
3	48	30	30	17	6
4	100	52	52	31	12
5	180	80	80	49	20
6	294	114	114	71	30
7	448	154	154	97	42
8	648	200	200	127	56
9	900	252	252	161	72
10	1210	310	310	199	90
11	1584	374	374	241	110
12	2028	444	444	287	132
13	2548	520	520	337	156
14	3150	602	602	391	182

Keďže metódy PH2 a PH3 sa navzájom líšia iba spôsobom kódovania tretej podmienky (každý výber vyberá iné alternatívy), tak obe kódovacie metódy pridávajú rovnaký počet pomocných symbolov.

IV. ZÁVER

Vykonané experimenty potvrdili, že zvyšovanie počtu použitých symbolov pri budovaní logických modelov sa paradoxne môže prejavovať na zmenšení veľkosti výsledného modelu. P-H metóda kódovania kardinalitných ohraničení patrí medzi takéto prípady, aj keď bolo nutné optimalizovať spôsob kódovania, pretože prvotný návrh priamo reflektujúci myšlienku výberov ďaleko zaostával za porovnávacou metódou binomického kódovania. Na druhej strane je potrebné upozorniť na prílišnú optimalizáciu ktorá pri našich experimentoch síce viedla k ďalšiemu zmenšeniu vytváraného modelu avšak doba spracovania tohto modelu sa naopak začala predlžovať.

LITERATÚRA

- [1] V. Kvasnička, *Matematická logika*, Bratislava, Slovenská Technická Univerzita, 2006
- [2] M. Mach, *Reprezentácia znalostí a riešenie úloh – logické prístupy*, Elfa, Košice, 2016.
- [3] E. Wynn, "A Comparison of encodings for cardinality constraints in a SAT solver", ArXiv, 2018.
- [4] S. Jabbour, L. Sais a Y. Salhi, "A Pigeon-Hole Based Encoding of Cardinality Constraints", Int. Symposium on Artificial Intelligence and Mathematics, Fort Lauderdale, FL, 2014.
- [5] S. Hattad et al., "Enhancing Pigeon-Hole based Encoding of Boolean Cardinality Constraints", 9th Int. Conference on Agents and Artificial intelligence, Scitepress, 2017, 299-307.
- [6] L. Rusnák, *Transformácia úloh s ohraničeniami na SAT problémy*, Diplomová práca, Fakulta elektrotechniky a informatiky Technická univerzita v Košiciach, 2021.

Secure routing in 6G network

¹Maros BAUMGARTNER, ²Natalia KURKINA, ³Jan PAPAJ,
⁴Richard NEZNIK

^{1,2,3,4}Department of Electronics and Multimedia Communications, Faculty of
Electrical Engineering and Informatics Technical University of Kosice, Slovak
Republic

¹maros.baumgartner@tuke.sk, ²natalia.kurkina@tuke.sk, ³jan.papaj@tuke.sk,
⁴richard.neznik@tuke.sk

Abstract – This article addresses the issue of secure routing as well as the issue of blockchain technology in 6G networks. It explains the role of one of the most prominent technologies to show its full potential systemically in 6G networks. It describes the challenge of eliminating problems for its deployment and solves the problem of decentralization through the use of Blockchain technology.

Keywords – 6G networks, blockchain, routing, security

I. INTRODUCTION

The recent boom in diversified mobile apps, especially those that support Artificial Intelligence (AI), have sparked heated debates about the future evolution of wireless communications. Fifth-generation mobile networks are already being deployed in parts of the world with the expectation of addressing the limitations of current cellular systems and providing an essential platform for intelligence services to evolve and thrive.

Today, we live in what can be described as a fully digital age, and its development is progressing steadily. Reliable data connectivity is essential for ever smarter, more automated technologies. Mobile data allows us to transmit quickly, and in a fully connected digital world we will need to connect literally everything, including people through vehicles, sensors, data to storage. Fifth-generation wireless networks that are currently being developed and deployed, offer significant advances beyond LTE. However, these networks may not be able to cover the fully interconnected requirements that will be necessary for our digital society.

II. 6G NETWORKS

Despite the long development of new technologies such as millimeter wave, massive multiple-input-multiple-output (MIMO) and the use of these higher frequency bands, it is now clear that 5th generation networks will not be able to meet all of these requirements to achieve the desired goal, even though 5G is vastly superior to its predecessors [2].

By 2030, our society is expected to move towards much more digitized, data-driven and intelligently inspired society that will be need near-instantaneous and ubiquitous wireless Internet connectivity. This will enable instant and unlimited wireless connectivity [9].

As a result, 6G is expected to improve the wireless technologies we know today, and achieve significantly higher system performance. So in the next few years several new applications will emerge to provide such integration. These should include, for example virtual and augmented reality, 8K video streaming, holograms, remote operations, Industry 4.0, smart homes, fog computing, integrated AI services or Unmanned Aerial Vehicles (UAVs), such as autonomous vehicles [3], [7].

Unlike previous generations, 6G will be transformative and revolutionize wireless evolution. Here we will move from connected things to connected intelligence that will have much more stringent requirements. These include very high data rates up to up to 1 Tb/s due to the sheer number of devices. Then there will need to be very high energy efficiency with the ability to support IoT devices without batteries, as well as massive low-response control (less than 1ms end-to-end) to achieve high speeds. It is also important to achieve very wide frequency bands (76 GHz - 140 GHz and 1 THz - 3 THz) to ensure high data rates. If we want to create broadband wireless network, it will be necessary to use not only the terrestrial network, but also wireless connectivity with satellite systems.

Hence, efforts from industry and academic field have focused on finding something that is more than 5G, and so the first 6G network concepts are emerging. In terms of speed, 6G is likely to use higher spectrum than previous generations with the aim of improving data rates, which are expected to be somewhere in the order of 100 to 1000 times faster than those of a 5G network [1].

In the frequency dimension, it will use a higher frequency spectrum than previous generations to improve transmission speeds. Specifically, 6G networks will enable connections at 100 Gb/s - 1 Tb/s. Examples include the combined use of the (1-3 GHz) band, the (30-300 GHz) millimeter wave band and the terahertz band (0,06 - 10 THz). On the other hand, in the future, mobile networks may be integrated with satellite systems and the Internet to form integrated space-to-ground networks, which will indeed increase the frequency ranges for services in terms of personal mobile communications [4], [5], [10].

A. Challenges for 6G network

The most important challenges in building a 6G network clearly include the following issues, that need to be addressed. These include:

- Massive connectivity in future systems
- Security requirements in future computing ecosystems
- Higher data consumption in sophisticated solutions
- Device resource limitations

Massive connectivity in future systems

Real time communication is a fundamental requirement for future ecosystems. Both device-to-device and machine-to-machine communication requires near-zero latency for the precision of its operations. Synchronization is a significant requirement in time-critical industrial applications. Critical backbone systems in countries, including power distribution systems and automotive networks, must be synchronized in real time for their operation to be accurate. Higher throughput is important for mission-critical systems that will leverage 5G and newer communication ecosystems requiring the simultaneous interconnection of a very large number of devices. This means that the network infrastructure as such should handle a huge volume of real-time transactions.

Security requirements in future computing ecosystems

Confidentiality is a very important tool for future infrastructures such as the IoT, which wireless connectivity exposes to huge threats. Encryption algorithms need to be simple for low-power IoT devices. Integrity is required when producing large amounts of data due to the fact that this data can be accessed by authorized users who will be able to modify it when it is transmitted again. Service availability is a fundamental requirement for future networks. In particular, the sophistication of 5G ecosystems with a large volume of interconnected devices will extend the risk of a DDoS attack. Authentication and access control is important for data that is transmitted or stored. Here we need to secure using control mechanisms and accesses to prevent tampering.

Audit is required to evaluate the behavior of tenants in the network ecosystem that use it. In the case of elevated security standards, packet-level auditing may be required to identify and flag the behavior of these tenants [3].

Higher data consumption in sophisticated solutions

Higher data rates are one of the most significant expectations in future network ecosystems. Applications such as virtual reality (VR), holographic communication, 16K video, and 3D ultra video require higher data rate and consumption.

Device resource limitations

Computational and storage constraints are expected to limit the capabilities of cryptographic algorithms and eventually lead to deviation from standard mechanisms. Standard security adoption is more difficult under such device resource limitations [3].

III. BLOCKCHAIN TECHNOLOGY

Blockchain is therefore seen as an essential technology to build trust for future networks. Because blockchain is considered to be one of the key enablers of technology for 6G mobile networks, it is also essential to explore the various benefits, opportunities and challenges that are expected in its use. The role of blockchain in 6G networks relates to the continuous creation of vertical services, differentiated services with networking. One of the important roles of blockchain and distributed ledger (DL) as a next generation distributed sensing services for 6G management applications [3].

Blockchain is composed of a certain number of blocks. Subsequently, the individual blocks are to form a blockchain. Each block contains a certain amount of data and is linked to the previous block. These blocks could be likened to a book, with the whole blockchain would form

a complete book together with the data that is on the pages. In addition to the data, each block contains a header consisting of some metadata about the block. Blocks are stored as files and are indexed off-string. For a blockchain to be useful, the data stored in it must be fixed and immutable. Hashing is used to be able to verify and at the same time ensure the integrity data [8].

Hashing

Hashing is the name for a process where, after input, we get a value in return. The output we get has a fixed length value, and its length is always the same no matter how much data it contains. This is a very long number, but in most cases it is represented using an alphanumeric value to make it easier to feel. Hashing is done using computer algorithm. If we enter the same data into the same algorithm, we always get the same output. If we change the data just a little bit, the resulting hash will be completely different. Even a small change, such as the use of lower or upper case letters, will affect the resulting hash value.

Nonce

This is a randomly generated number. The reason for generating this value is to add a different value to this number. This value gives us a different resulting hash. We can generate different hashes by changing this generated number [8].

IV. RESULTS

As the simulations showed, reliable selection of transmission nodes was solved by quality of service and quality of experience parameters. Node selection using the proposed algorithms showed how reliable packet transmission could be addressed in a 6G network. In the initial search for reliable transmission, a threshold was reached of finding such a transmission at 60%. Another fact was the real number of packets transmitted compared to the simulation. If we take into account packet transmission in the simulation, we achieved a 69.2% transmission success rate. This scenario represents packet transmission in the simulated environment. If we had the same transmission in a real environment, it would be necessary to incorporate interference from nearby sources.

After adding additional users, the number of paths found that are suitable for further data transfer from a reliability point of view has increased. It is due to the addition of the new users that the value of finding a reliable transfer has increased to 91%. The addition of users alone also made the whole simulation more challenging in terms of processor load, as the number of combinations needed to execute the whole algorithm was obviously multiplied. When selecting the appropriate nodes for further data transmission, this selection was immediately governed only by parameters such as the number of packets lost, quality of experience. These two parameters form the backbone of the entire selection of suitable nodes for further packet routing in terms of reliability. In terms of packets transmitted, we achieved 77% transmission simulation results. If we take into consideration the interference, the real value of transmitted packets would be 69.4%.

Game theory is used to select suitable nodes. The input parameters are QoS and QoE, the output is a mixed strategy that is obtained from the two parameters. In terms of secure routing, the node that has the best value of this mixed strategy for a specific link is important. Parameters such as number of lost packets and quality of experience were used in this selection. These are QoS and QoE parameters, which formed the basis for selecting suitable nodes for reliable transmission from the security point of view.

At figure 1 (a) we can see the difference between reliable path for 30 and 50 users. The green color indicates a previously found reliable path for 30 users. Red color indicates a reliable path after adding additional new users to the process of selection. If the new users are added, the whole path can be different. there has also been an increase in suitable paths found after adding newusers to our network.

Figure 1 (b) shows the percentage of reliable path transfer for 50 users. This graph represents the transmitted packets in simulation and real transmitted packets. We can see the difference in the simulation compared to the real transmission, where we are considering interference. The difference between simulation and real transmission is about 10%. At figure 3 we can see the creating of blockchain. Each block contains index, some data, previous hash in case it's not genesis block, self-hash and nonce. Each block contains data as current user, his neighbor, number of packet loss, quality of experience and time stamp. The last figure 4 shows how the additional users are connected to the existing blockchain. After adding new users to the blockchain we have a onebig blockchain which contains every user, who is suitable for the future consideration in selection in terms of security.

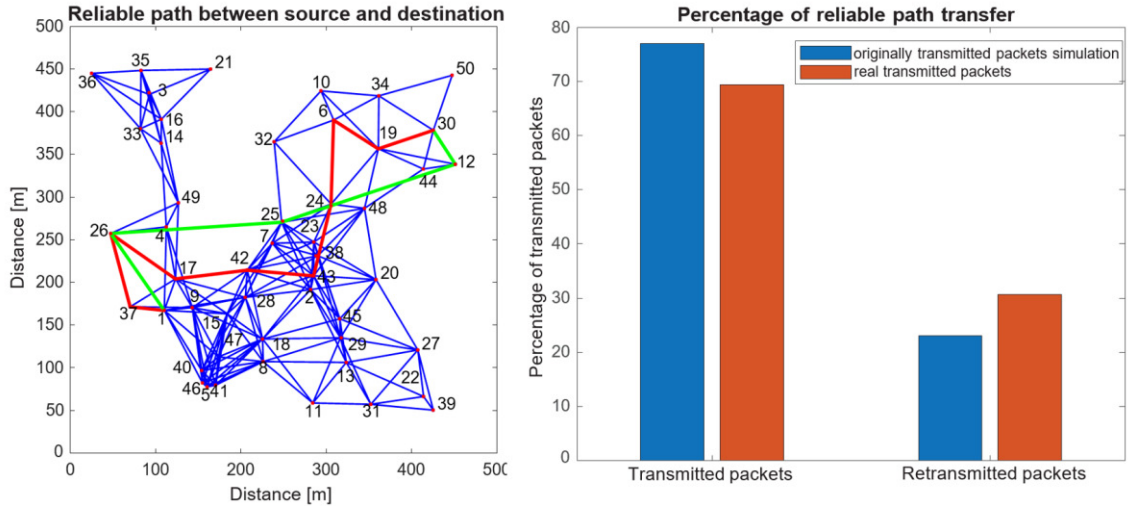


Fig. 1 Reliable path between source and destination for 30 and 50 users (a) Percentage of reliable path transfer for 50 users (b)

```

newBlock =
Block with properties:
    index: 1
    data: {'1' '26' '2' '87' '12-Apr-2022 21:58:48' 'Bezpečnostné_hodnotenie'}
    previousHash: '075c27741a3506846368fa6e5b3477f95b31ccee71a5716e2f12b40fa21d23aa'
    selfHash: []
    nonce: 112

newBlock =
Block with properties:
    index: 1
    data: {'1' '26' '2' '87' '12-Apr-2022 21:58:48' 'Bezpečnostné_hodnotenie'}
    previousHash: '075c27741a3506846368fa6e5b3477f95b31ccee71a5716e2f12b40fa21d23aa'
    selfHash: '00df8b6481d2b0cde1e2dd599286ab5'
    nonce: 112

newBlock =
Block with properties:
    index: 2
    data: {'26' '25' '2' '87' '12-Apr-2022 21:58:49' 'Bezpečnostné_hodnotenie'}
    previousHash: '00df8b6481d2b0cde1e2dd599286ab5'
    selfHash: []
    nonce: 233

newBlock =
Block with properties:
    index: 2
    data: {'26' '25' '2' '87' '12-Apr-2022 21:58:49' 'Bezpečnostné_hodnotenie'}
    previousHash: '00ac70646a12ccfec4b3af5823e37bd'
    selfHash: []
    nonce: 233

Block with properties:
    index: 5
    data: {'101' '101' '101' '101' '12-Apr-2022 22:03:54' 'Bezpečnostné_hodnotenie'}
    previousHash: '004f8fbb41c915bf483373c918d167c8'
    selfHash: []
    nonce: 148

newBlock =
Block with properties:
    index: 5
    data: {'101' '101' '101' '101' '12-Apr-2022 22:03:54' 'Bezpečnostné_hodnotenie'}
    previousHash: '004f8fbb41c915bf483373c918d167c8'
    selfHash: '00e2fa2a24a26cd62a52f11d6bea6bf'
    nonce: 148

>> vytvaranie_blockchain pridane_uzly

newBlock =
Block with properties:
    index: 6
    data: {'1' '37' '2' '18' '12-Apr-2022 22:04:03' 'Bezpečnostné_hodnotenie'}
    previousHash: '00e2fa2a24a26cd62a52f11d6bea6bf'
    selfHash: []
    nonce: 334

newBlock =
Block with properties:
    index: 6
    data: {'1' '37' '2' '18' '12-Apr-2022 22:04:03' 'Bezpečnostné_hodnotenie'}
    previousHash: '00e2fa2a24a26cd62a52f11d6bea6bf'
    selfHash: '00c70acd35c33ffa5fb7606d2d9e04cf'
    nonce: 334
    
```

Fig. 2 Creating blockchain for suitable users (a) Connecting new users to an existing blockchain (b)

V. CONCLUSION

Security in terms of routing in 6G network is still really actual topic. The direction in which today's society is heading contributes with the drive to connect more and more devices to the internet. Security issues are important today, and the future won't be different. Concepts such as autonomous driving are slowly coming into mainstream use and therefore it is extremely important that we do not forget the other side of the matter, which is safety.

Thanks to blockchain technology, we have been able to increase the overall reliability of the transfer. By adding blockchain, we ensure that it is much more difficult to modify, hack or steal some data. Since the blockchain itself uses a hashing function for each block itself, this way it increases the overall security of the transmission. We can also use the proposed algorithm tell which nodes are suitable for further packet transmission from the reliability point of view.

The simulations showed that the overall selection is greatly influenced by the generated topology. A reliable path will always be generated, provided the conditions for its implementation are met. One case of failed path generation may be that a topology is generated where the destination node is not within range of the with any other node. With a very high number of devices, it should be assumed that such a simulation may take several hours or days.

ACKNOWLEDGMENT

This publication has been published with the support of the Operational Program Integrated Infrastructure within project: Research in the SANET network and possibilities of its further use and development, code ITMS: NFP313010W988, co-financed by the ERDF."

REFERENCES

- [1] CHEN, Huangqing, et al. 6G Wireless Communications: Security Technologies and Research Challenges. In: 2020 International Conference on Urban Engineering and Management Science (ICUEMS). IEEE, 2020. s. 592-595
- [2] DAVID, Klaus; BERNDT, Hendrik. 6G vision and requirements: Is there any need for beyond 5G?. IEEE vehicular technology magazine, 2018, 13.3: s. 72-80.
- [3] HEWA, Tharaka, et al. The role of blockchain in 6G: challenges, opportunities and research directions. In: 2020 2nd 6G Wireless Summit (6G SUMMIT). IEEE, 2020. p. 1-5.
- [4] HENRIQUE, Paulo Sergio Rufino; PRASAD, Ramjee. 6G The Road to the Future Wireless Technologies 2030. River Publishers, 2021.
- [5] MAIER, Martin; EBRAHIMZADEH, Amin. Toward 6G: A New Era of Convergence. John Wiley and Sons, 2021.
- [6] WANG, Minghao, et al. Security and privacy in 6G networks: New areas and new challenges. Digital Communications and Networks, 2020, 6.3: 281-291.
- [7] RONG, Bo. 6G: The Next Horizon: From Connected People and Things to Connected Intelligence. IEEE Wireless Communications, 2021, 28.5: s. 8-8.
- [8] SUMMERS, Akira. Understanding Blockchain and Cryptocurrencies: A Primer for Implementing and Developing Blockchain Projects. CRC Press, 2022.
- [9] XU, Hao, et al. Blockchain-enabled resource management and sharing for 6G communications. Digital Communications and Networks, 2020, 6.3: 261-269.
- [10] YANG, Ping, et al. 6G wireless communications: Vision and potential techniques. IEEE network, 2019, 33.4: 70-75.

Visual Interaction with Intelligent Space

Ján VAŠČÁK

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

jan.vascak@tuke.sk

Abstract – This paper presents an overview of the issue of visual interaction with smart space environments. It also provides an overview of the technologies used in this interaction that enhance the level of interaction between the user and the system. These technologies include the visualization itself, voice control, and system autonomy. In implementing this system, we have created a web application incorporating the aforementioned technologies. The result is an application freely accessible via an Internet connection.

Keywords – cloud systems, intelligent space, Raspberry Pi, visualization, voice assistants

I. INTRODUCTION

Nowadays, it's becoming more and more common to control the devices around a person using software designed specifically to control them. Whether it's common household appliances or more robust systems for a larger set of devices. The aim of this thesis is to point out common software and hardware means with which it is relatively easy to realize a visual interaction with the smart space environment, represented by both a 3D model and a control module, um allowing either manual control or automatic control using an expert system. For this purpose, both a brief overview of the state-of-the-art resources used for this type of tasks and the actual design of the architecture of the process control solution in the school laboratory are presented. In the following, we show the functioning of the application, visualization and control methods. Finally, we will point out possible future improvements and extensions.

II. MEANS OF INTERACTION WITH A SMART SPACE ENVIRONMENT

Here, we give a brief overview of the basic components that are used in the design of systems for remote human-environment interaction, such as means of environment visualization, machine learning, voice control, or cloud services.

The main goal of visualization in systems is to simplify the representation of the main parts of the system. Initially, there was only 2D visualization, which was very successful, and for this reason, during the development of computing, 3D visualization came into being with new possibilities [1].

Experiences during software development using visualization record a better understanding of the system with lower development costs. The essence of visualization is the creation of an image of a given system using visualized objects. Often these objects are visualized while running certain processes in a more complex system and also 3D objects containing animations are included, which increases the attractiveness of the visualization system and the memorability of the human visual system. State-of-the-art 3D modelling tools include ArchiCAD, Tekla and Blender [1].

In this work, we are most concerned with the visualization of smart spaces. The goal was to create a 3D model representing an intelligent space capable of interacting with the user. An essential part in the creation of such a model is the optimization due to the 3D animations played in real time. A fundamental characteristic of visualization systems is to maintain a high frame rate per second. The average value is 15Hz, and if a lower frequency occurs, the interest of the user's interaction with the system decreases. In the rendered space, there is a limit to the number of 3D objects that can be optimized using various techniques [1]:

- 1) Optimization – takes advantage of the available hardware and uses the GPU to perform the highest quality requests. In a situation where GPU performance is not sufficient, the system uses the CPU. Visualization can be optimized by using objects pre-stored in GPU memory, thus no data transfer is required for individual frames. Other ways of optimization are combining geometric data of objects located next to each other, reducing the quality of object materials, and rendering the same objects together.

- 2) Level of detail – is a proportional decrease in the quality of objects depending on the distance and the current view. This means that when more distant objects are displayed with lower quality, there is no lower visual impression for the user.
- 3) Object Visibility – represents the display of objects located in the visible area of the camera. Objects located outside this view do not need to be rendered, for this reason the performance is used more for visible objects. This principle is a common component when visualizing large spaces.

Modern technologies are often part of smart spaces, such as buildings, laboratories or ordinary households, which are required to be somewhat autonomous. Machine learning and artificial intelligence algorithms are therefore often used for these purposes [2].

One of the means used for this purpose are expert systems, which are based on rules that are close to human knowledge representation. An example of a rule in conjunction with an intelligent space looks like the following [2]:

IF the temperature ≤ 20 THEN turn on the radiator ELSE turn off the radiator. (1)

A set of such rules is called a "rule-based system". The basis is a set of rules, where each rule represents a condition containing the so-called precondition (left) and conclusion (right) sides of the context rule. These rules are defined by an expert who can react correctly to specific situations. The left-hand side of the rule consists of several conditions that are compared with their facts, all of which must be satisfied. When several rules are applied, their conclusions are merged and the degree of influence of a given rule on the overall output depends on its so-called strength, which is the degree of compatibility of its assumption with the real measured inputs [2].

Voice control greatly enhances the user's level of interaction with the smart space. Voice interaction with systems is one of the important goals of artificial intelligence. Also, this field is one of the fastest developing. Examples of these systems are the voice assistants Cortana from Microsoft, Siri from Apple, Alexa from Amazon or Assistant from Google. Their main task is to solve various tasks quickly and efficiently through spoken speech. Nowadays, these systems are not limited by the type of questions, but can respond to a wide range of requests. The design method depends on the application and its complexity [3].

In the future, voice assistants are expected to be a common feature of phones, smart glasses, kitchen appliances, televisions, game consoles, and other similar devices. These systems form an important part of human interaction with the smart space. The main reason for this is the unnecessary of manual interaction, which leads to an increase in user comfort [3].

The actual design of such a system depends on the complexity of the application. The most complex design is the so-called multi-modal dialog system, which offers the processing of multiple combinations of user inputs such as speech, images or gestures [3].

Cloud services are a frequent part of voice assistants in particular. The most well-known services include:

- 1) LUIS – a service from Microsoft that can detect the meaning of a sentence using machine learning technologies and then process the key information obtained [4].
- 2) QaA Maker – also a service from Microsoft, which is used as a knowledge base containing static information if the same type of answer is always sufficient [5].
- 3) Speech-to-Text – Google's service that can efficiently transform a spoken sentence into text [6].
- 4) Text-to-Speech – also a Google service that transforms text into human speech [7].

III. ARCHITECTURE OF VISUAL INTERACTION WITH INTELLIGENT ENVIRONMENT

The goal was to make the application easily accessible to multiple users. For this reason, we decided to develop a web application consisting of a front-end and two application programming interfaces (APIs). The first API runs on a Raspberry Pi 3 Model B device, to which the GrovePi hardware interface with AllThingsTalk sensors is connected. The API programmed in the Express library is used for voice control purposes, which will be described in more detail later. Communication between the components is done using a REST architecture. We chose PostgreSQL as the database, which is also running on the Raspberry Pi. Fig. 1 shows the overall architecture of the proposed solution.

The primary goal was to create a 3D model of the space. For the realization of this problem we chose Blender 4.0 software. We visualized the laboratory of the Internet of Things, Department of Cybernetics and Artificial Intelligence, Technical University of Košice, see Fig. 2.

On the front-end, we used the React JavaScript library, which can import a 3D model created by Blender using the ThreeJS library. We imported the model using GLTF Loader, which efficiently

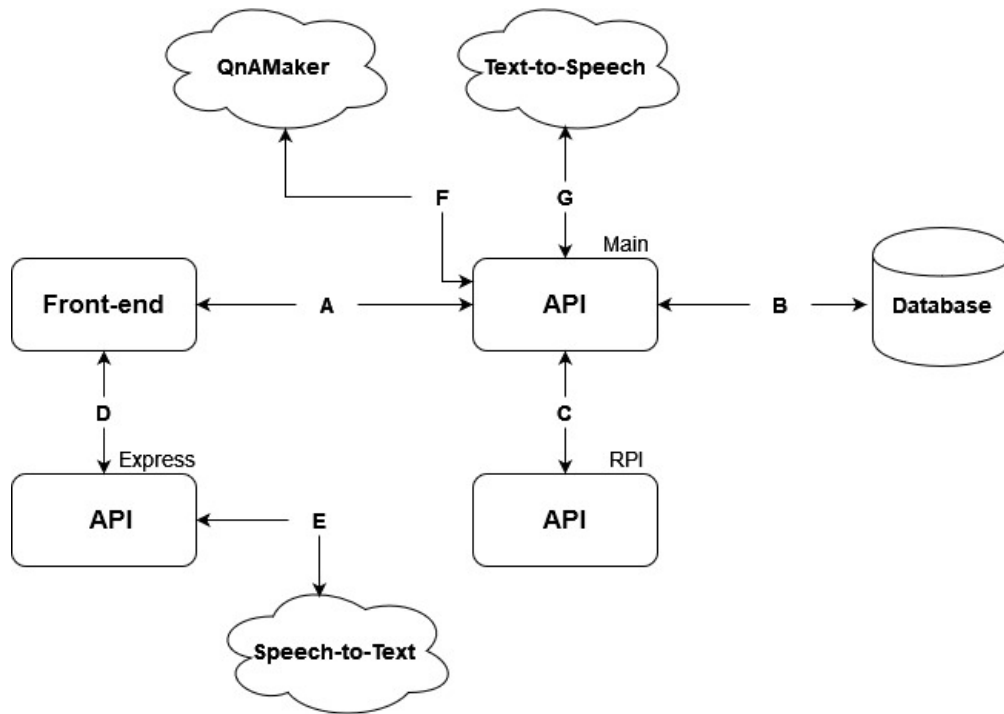


Fig. 1 Architecture of the application for remote interaction with the intelligent environment.

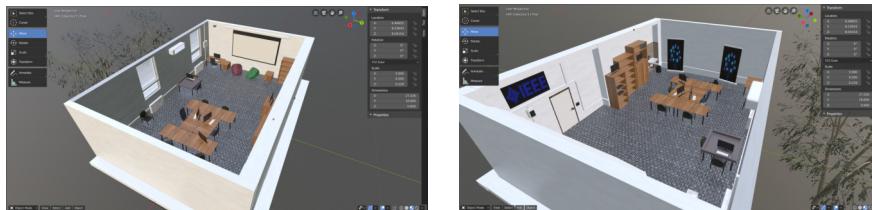


Fig. 2 Two visualization views of the used smart space.

renders the model with its materials. The application also simulates the adverse data from the gas and fire sensors. If these values are recorded, the visualizations are turned on, see Fig. 3. Other visualized devices include conditioning in the form of 3D object color change, window blinds rotation, and real-time camera streaming. The system also recorded movement in the space, temperature, and humidity.

As seen in Fig. 1, we have created 2 APIs. The first, running on a Raspberry Pi, is programmed using Python and the Flask library. The main role of this API is to provide real sensor data to the client application. The following sensors were connected to the GrovePi module:

- 1) DHT11 – temperature and humidity sensor
- 2) PIR – motion detection sensor
- 3) Magnetic switch – sensor simulates door opening
- 4) MQ3 – gas sensing sensor
- 5) KY-026 – fire detection sensor
- 6) Switch – simulates switching on the air conditioning
- 7) Light sensor – offers better visualization of the lighting in the space
- 8) USB Webcam – offers real-time streaming
- 9) Rotor – simulates the rotation of window blinds
- 10) 3 LEDs – simulates light on in the room
- 11) Buzzer – simulates an alarm in case of fire or gas detection

This API is connected to the QnAMaker cloud service for the need of voice control. It also offers autonomous space management. The second API, programmed in JavaScript using the Express library, was created because of the use of the cloud-based Text-to-Speech and Speech-to-Text services. The Flask library was unable to properly handle the audio data sent from the client application, and at the same time, these services were unable to accept requests due to running the API under the Evelnet package. Express used these services seamlessly.

Table 1
 Example of expert system rules.

Time	Lights	Movement	Gas	Fire	Windows	Buzzer	Air	Temperature	Action
>16	no	yes							Switch lights on.
						no		yes	Close windows.

Voice control provides interaction using the user’s voice. This voice assistant can be used to detect the status of the device by voice or to switch it to another state. Due to the complexity of developing such an assistant, we decided to use cloud services.

The application consists of a dashboard for using this service. There are two buttons in the panel and the bottom part shows the text based on the voice in real time. When the "Record" button is pressed, the voice recording is started and sent to the Speech-to-Text service. When the "Stop" button is pressed, the recording stops and the entire sentence is sent to QnAMaker, which evaluates it and sends back the appropriate response. This response is then sent to the front-end application, which generates a full-sentence response, which then enters the Text-to-Speech service, which is used to trigger a voice response from the system.

This system has been programmed using Python and the Expert library for expert system development and runs on a Raspberry Pi device. We created 11 rules that define the behavior of the system based on the recorded sensor data. V Table 1 are two of the eleven expert system rules. The header contains the inputs to this system, based on which an action is inferred. This system is triggered if the user chooses to enable smart mode in the application.

Once the overall web application was created, the goal was to make it accessible to more users. We ran the front-end application on Heroku’s cloud server. In the same way, we handled the deployment of the API used for voice control. We handled remote access for the API running on the Raspberry Pi using an external IP address and port forwarding on the local network.

IV. EXPERIMENTS

The experiments themselves consisted of testing the functionality and speed of the system’s response to changes in environmental states and commands from a remote user. Model situations were set up. In the following, we present one of them for illustration.

In Fig. 3 the visualization window shows the initial state of the environment, namely the opening of both windows and the turning on of the light. When a command is entered, changes in the model are reflected in real time using socket communication.

The aim of this experiment is to return the original state using a 3D model. Clicking on both windows and the light switch will return the state to the original one, indicating the successful completion of the experiment. The given action is also signaled by the control LEDs that are connected to the Raspberry Pi and this way the user also receives information about the execution of the action.

Similarly, we can test a safety system consisting of fire and gas simulation in a 3D model. For example, if a fire is detected in a room, the visualization in the 3D model is also triggered, see Fig. 4a. Similarly, the behavior of other sensors and devices is also visualized. E.g. in the visualization, the movement of motors is recorded, which on command close windows (Fig. 4b), or switching on the air conditioner causes the color of its icon to change in the 3D model.

Fig. 5 shows a voice control panel with the recorded sentence, "Close window one." This command causes the model to close the window closer to the canvas wall. Other devices and sensors can be referenced in a similar manner.

In the basic control window it is possible to switch the system to the so-called smart mode (button "Smart mode"), in which the expert system is started and the required actions are performed based on the defined rules, see Table 1.

V. CONCLUSIONS

The system created for visual interaction with the smart space works reliably even when multiple users are controlling it at the same time. A limitation is the absence of security control on the API running on the Raspberry Pi. A solution would be to link the Raspberry to a server that provides this certificate. As an improvement to the application, a better recording system of the user’s voice and a more accurate method for recognizing the meaning of a sentence would be useful. The expert system also works reliably on the current state of the sensors in the space. Another improvement would be the use of a better camera for real-time streaming needs.

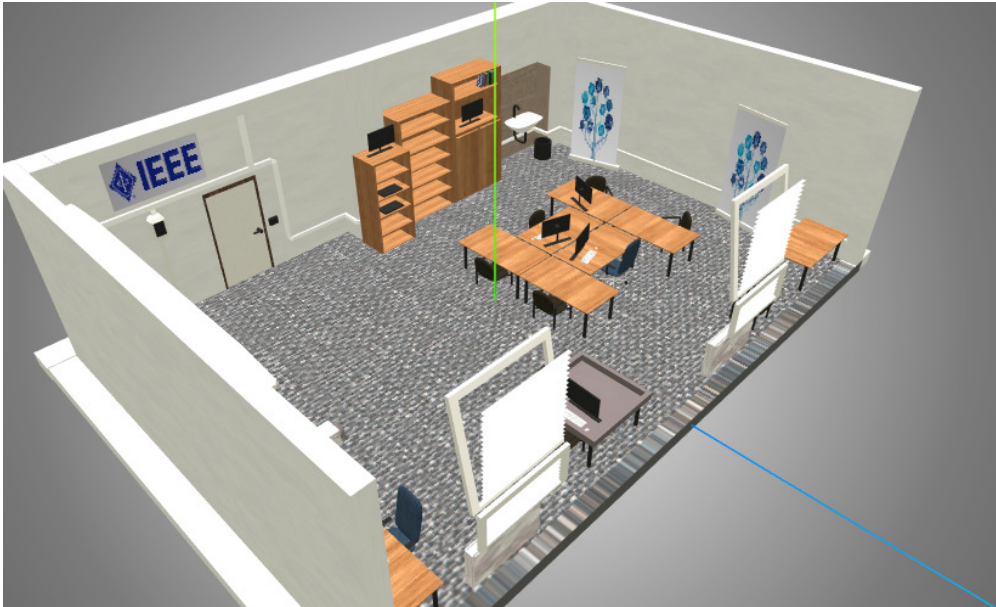


Fig. 3 Input visualization.



Fig. 4 Examples of actions in a 3D visualization model: a) – fire animation; b) – window closing animation.



Fig. 5 Voice Control Panel (bottom right window).

ACKNOWLEDGMENT

This publication is the result of the APVV grant ENISaC – Edge-Enabled Intelligent Sensing and Computing (APVV-20-0247).

REFERENCES

- [1] M. Johansson, M. Roupé, and P. Bosch-Sijtsema, "Real-time visualization of building information models (BIM)," *Automation in Construction*, vol. 54, pp. 69–82, 2015.
- [2] I. Uddin, "A rule-based framework for developing context-aware systems for smart spaces," July 2019. [Online]. Available: <http://eprints.nottingham.ac.uk/57019/>
- [3] V. Kěpuska and G. Bohouta, "Next-generation of virtual personal assistants (Microsoft Cortana, Apple Siri, Amazon Alexa and Google Home)," in *2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC)*, 2018, pp. 99–103.
- [4] "What is LUIS?" 2022. [Online]. Available: <https://docs.microsoft.com/en-us/azure/cognitive-services/luis/what-is-luis>
- [5] "From data to bot in minutes," 2022. [Online]. Available: <https://www.qnamaker.ai/>
- [6] "Speech-to-text," 2022. [Online]. Available: <https://cloud.google.com/speech-to-text>
- [7] "Text-to-speech," 2022. [Online]. Available: <https://cloud.google.com/text-to-speech/>

Lineárny pohon

¹Peter GIROVSKÝ, ²Tadeáš KMECIK

^{1,2} Katedra elektrotechniky a mechatroniky, Fakulta elektrotechniky a informatiky,
Technická univerzita v Košiciach

¹peter.girovsky@tuke.sk, ²tadeas.kmecik@student.tuke.sk

Abstrakt — V tomto príspevku je prezentovaný návrh hardvérovej časti lineárneho pohonu pre univerzálne použitie v rámci rôznych zariadení. V práci je predstavený návrh mechanickej a elektrickej časti pohonu ako aj praktické overenie funkčnosti elektrickej časti lineárneho pohonu.

Kľúčové slová — lineárny pohon, prevodovka, jednosmerný motor, riadenie

I. ÚVOD

V súčasnosti ide technika vpred veľmi rýchlo a vyžaduje sa čím ďalej tým vyššia presnosť, rýchlosť a kvalita výroby. Tieto vlastnosti sú v značnej miere v mnohých činnostiach ovplyvnené ľudským faktorom. Preto sa v súčasnej dobe snažíme redukovať ľudský faktor. Aj keď existujú činnosti pri ktorých je človek nenahraditeľný, tak aj tam je snaha čo najviac uľahčiť a redukovať úkoly človeka. Pri cyklických činnostiach vykonávaných človekom, použitím vhodných pohonov mnohonásobne zvyšujeme produkciu. Preto vývoj a riešenie problematiky pohonov je v praxi nesmierne dôležité. Pohony, ktoré sa používajú na rýchle a presné riadenie polohy, nazývame servopohony alebo tiež aktuátory. Tieto pohony sa skladajú z motora, prevodovky, snímača polohy a v prípade potreby zmenu zmyslu energie obsahujú aj konkrétny typ prevodov. Najčastejšie používaným pohonom v praxi je elektrický lineárny aktuátor, ktorý prevádza rotačnú mechanickú energiu motora, najčastejšie jednosmerného motora, na translačný priamočiary pohyb pomocou vhodného druhu prevodu. Takýto lineárny aktuátor má praxi časté využitie najmä tam kde sa vyžaduje presné riadenie polohy, manipulovanie s ťažkými objektami, kde daný objekt je buď ťahaný alebo tlačný. Riadenie týchto pohonov najčastejšie zabezpečuje mikrokontrolér s nahratým programom riadenia. Keďže dnešná doba si vyžaduje veľmi rýchly návrh a vývoj takýchto pohonov, preto sa tieto aktuátory nenavrhnú pre konkrétne použitie v danom pracovnom mechanizme ale ako univerzálne moduly s rôznymi parametrami pre všestranné použitie v praxi. Takýto prístup vývoja pohonov umožňuje redukovať cenu a samotný čas potrebný pre realizovanie funkčných mechanizmov tvorených týmito elektrickými aktuátormi. Veľká popularita týchto elektrických pohonov je aj kvôli ich výhodným vlastnostiam, ktorými sú napríklad bezpečná a relatívne čistá, tichá prevádzka, jednoduchosť riadenia, ľahkosť dosiahnutia veľmi presného polohovania a kontroly polohy. Ďalšie výhodné vlastnosti týchto pohonov sú napríklad ich elektrická a mechanická účinnosť, veľmi dlhá životnosť, nízke nároky na údržbu a vykonávanie servisu. Používanie elektrických aktuátorov je v mnohých oblastiach výhodnejšie ako použitie pneumatických alebo hydraulických pohonov. To hlavne kvôli ich kompaktnosti a nepotreby použitia veľkých zdrojov energie akými sú zásobníky stlačeného vzduchu, vysokotlakové čerpadlá s nádobou na hydraulickú kvapalinu. Presne kvôli týmto popísaným výhodám je použitie elektrických aktuátorov v praxi ideálnou voľbou pri riešení pohonu pre rôzne praktické použitia v skoro všetkých mechanických zariadeniach.

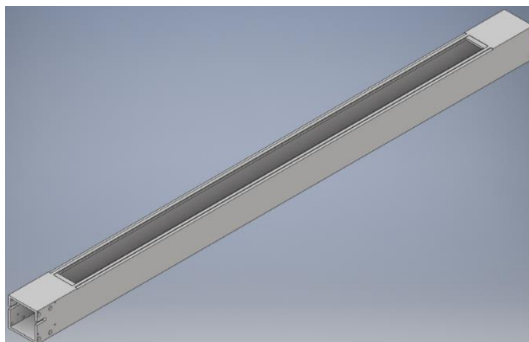
II. NÁVRH MODELU LINEÁRNEHO POHONU

Návrh samotného hardvéru modelu lineárneho pohonu sa bude skladať z mechanickej a elektrickej časti. Mechanická časť pozostáva z návrhu puzdra vedenia, puzdra pre ozubené remenice s možnosťou napínania ozubeného remeňa, držiakov lineárneho vedenia, puzdra pre neodýmové magnety a vonkajšieho polohovacieho vozíka. Pri návrhu mechanickej časti je nutné uvažovať aj potrebu upevnenie inkrementálneho snímača a takisto snímačov koncovej polohy.

A. Mechanická časť pohonu

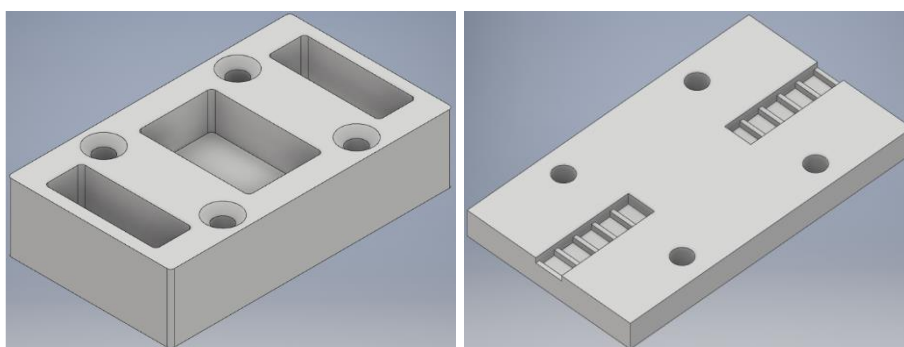
Puzdro vedenia - táto časť lineárneho pohonu zobrazená na obr.1, slúži ako hlavná nosná

konštrukcia na ktorú sa montujú ďalšie mechanické komponenty. Základ tohto puzdra vychádza z duralového profilu rozmeru 75x75x5, ktorý svojimi vlastnosťami nám poskytuje dostatočnú pevnosť a robustnosť pri čo najmenšej hmotnosti.



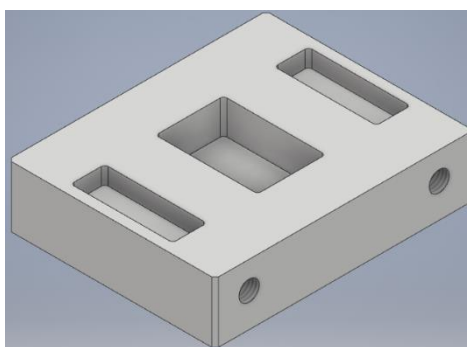
Obr. 1 Puzdro vedenia

Puzdro magnetov – táto časť slúži na osadenie vnútorných neodýmových magnetov, pomocou ktorých sa bude prenášať energia cez mosadznú planžetu na polohovací vozík. Na puzdre, ktoré je uvedené na obr. 2 môžeme vidieť vyfrézované otvory do ktorých sa osadia neodýmové magnety. Parametre zvolených magnetov sú KV 30x20x10 mm materiálu N42 s odtrhovou silou 19 kg, ktoré budú umiestnené v strede vozíka a na jeho krajoch budú umiestnené magnety KV 30x10x10 mm takisto materiálu N42 s odtrhovou silou 10 kg. Odtrhová sila magnetov je nadimenzovaná tak, aby kompenzovala straty, ktoré vznikajú vo vzduchovej medzere.



Obr. 2 Puzdro magnetov a spojky remeňa

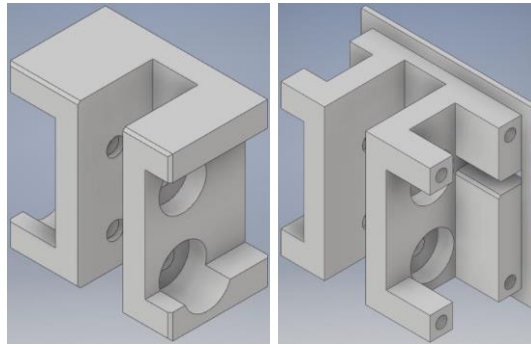
Polohovací vozík – zobrazený na obr.3 je vyrobený z duralového materiálu ktorý zabezpečí vysokú pevnosť a odolnosť pri nízkej požadovanej hmotnosti, ktorá je dôležitá kvôli dosiahnutiu čo najnižšej zotrvačnosti pohyblivých častí.



Obr. 3 Polohovací vozík

Držiaky ozubených remení – lineárny pohon pozostáva z dvoch držiakov ozubených remení, zobrazených na obr.4., pričom jeden z nich bude pevne spojený s puzdrom vedenia a

druhý sa bude pohybovať vo vnútri vedenia, nakoľko pomocou neho sa bude napínať ozubený remeň.



Obr. 4 Držiaky ozubených remení

Ostatné komponenty – samotná mechanická časť pohonu pozostáva, okrem už popísaných hlavných komponentov, aj z ďalších dôležitých častí bez ktorých by pohon nefungoval ako napr. hriadele, remenice, remene, spojka spájajúca hlavný hnací hriadeľ s hriadeľom enkodéra, koncové kryty a pod.

B. Elektrická časť pohonu

Elektromotor – pre lineárny pohon bol ako aktuátor zvolený jednosmerný cudzobudený motor s permanentnými magnetmi MAXON RE 35 s nasledujúcimi parametrami:

- Nominálne napätie 15V
- Nominálny prúd 4A
- Nominálny moment 0,0742 Nm
- Nominálne otáčky 6500 ot/min
- Účinnosť 85%

Snímač polohy - ak chceme jednosmerný motor riadiť ako servopohon, potrebujeme poznať informáciu o aktuálnej polohe rotora. V našom prípade pre snímanie polohy bol použitý inkrementálny snímač ISC3806. Parametre inkrementálneho snímača:

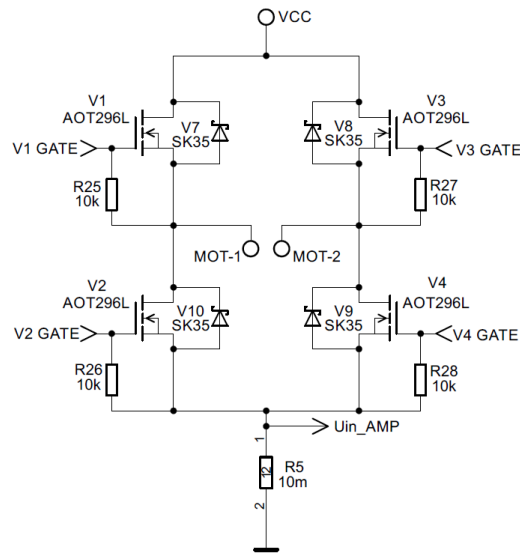
- Napájacie napätie 5-12 V
- Maximálne otáčky 5000 ot/min
- Počet generovaných impulzov jednej fázy na otáčku 1000 imp/ot

Snímače koncových polôh - snímač koncovej polohy funguje ako nadradená ochrana, v prípade zlyhania niektorej elektrickej časti. Ak polohovací vozík nadíde na niektorú z krajných polôh okamžite sa zablokuje činnosť budičov MOSFET tranzistorov a tým sa vypne napájanie motora. Parametre vybraného indukčného snímača **NPN/NC**:

- Napájacie napätie 5-36 V DC
- Detekčná vzdialenosť 0-8 mm
- Maximálny spínaný prúd 300 mA.

Menič pre jednosmerný motor – jednosmerný motor s konštantným budením (budenie pomocou permanentných magnetov) je možné riadiť iba zmenou napájacieho napätia a smer otáčania zmenou polaritý napájacieho napätia. Navrhnutý menič pre napájanie jednosmerného motora pozostáva z niekoľkých častí:

- Výkonová časť – H mostík, zobrazený na obr. 5, je tvorený MOSFET tranzistormi **AOT296L**
- Obvody budenia MOSFET tranzistorov výkonovej časti – budiče **IR2110PBF**
- Obvod brzdenia jednosmerného motora
- Oddelovacie obvody slúžiace na galvanické oddelenie riadiacej a výkonovej časti
- Obvody pre snímanie prúdu a napätia
- Napájací obvod – impulzný menič **AP1501-12**



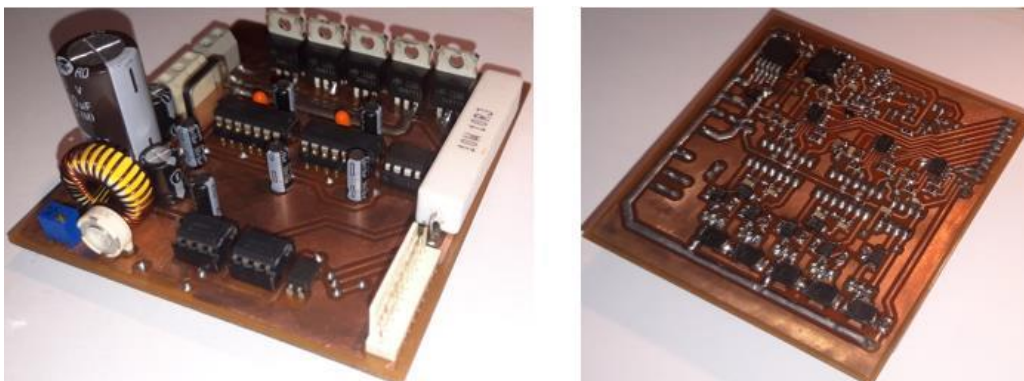
Obr. 5 Schéma H-mostíka

III. PRAKTICKÁ REALIZÁCIA NÁVRHU LINEÁRNEHO MOTORA

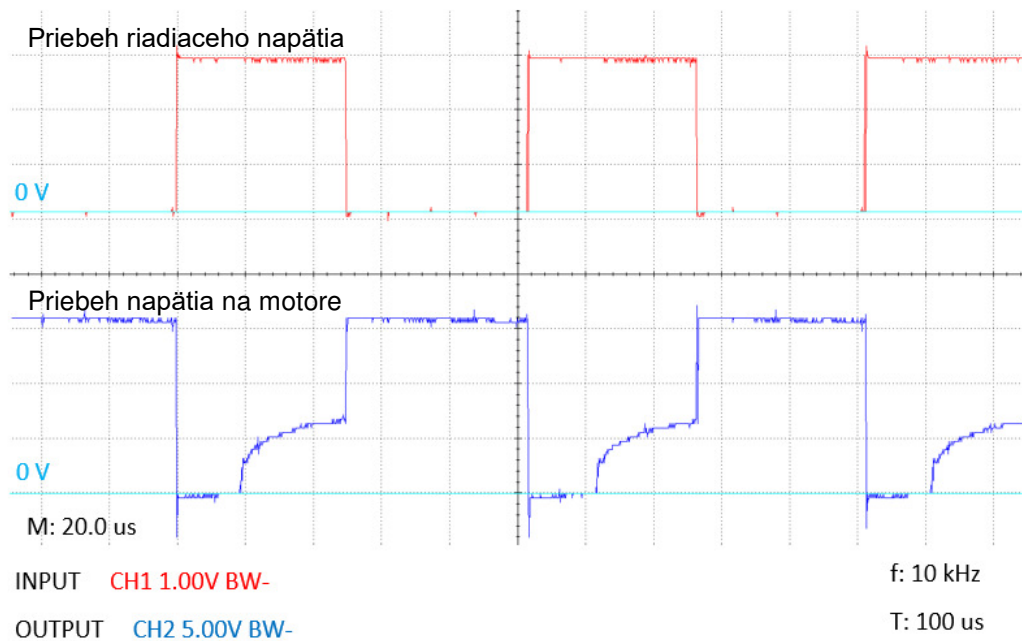
Na obr. 6 môžeme vidieť celkovú mechanickú časť pohonu už s osadenými elektrickými komponentami. Táto časť sa skladá z hlavného puzdra vedenia, do ktorého sa pomocou držiakov na spodnú časť pripevnilo prizmatické vedenie spolu s vozíkom. Medzi prizmatický vozík a puzdro na magnety bola vložená spojka remeňa, ktorá slúži na mechanické prepojenie ozubeného remeňa a do puzdra vedenia boli zo strán vložené držiaky, už s osadenými remenicami a hriadeľmi. Z vonkajšej strany je upevnený inkrementálny snímač, snímače koncových polôh a motor. Navrhnuté elektronické obvody sú zobrazené na obr. 7.



Obr. 6 Mechanická časť lineárneho pohonu so zabudovaným jednosmerným motorom a inkrementálnym snímačom



Obr. 7 DSP s elektronickými obvodmi



Obr. 8 Priebeh PWM signálu z meniča spolu s priebehom napätia na motore

V rámci testovania bola odskúšaná základná funkčnosť elektrickej časti lineárneho pohonu a tým sa odskúšala aj základná funkčnosť toto pohonu. Na obr. 8 je červenou farbou zobrazený priebeh predstavujúci riadiace napätia z mikrokontroléra, ktorý generuje signál PWM so striedou 0,5 a modrý priebeh predstavuje napätie na jednosmernom motore s permanentnými magnetmi.

IV. ZÁVER

V tomto príspevku sme predstavili návrh konštrukcie lineárneho motora. Návrh konštrukcie pozostával z dvoch základných častí, mechanickej a elektrickej, pričom výsledkom práce je kompletne realizovaná hardvérová časť lineárneho pohonu, ktorej funkčnosť bola aj overená. Realizovaný lineárny pohon je univerzálne použiteľný v rôznych aplikáciách, kde sa vyžaduje presný lineárny pohyb, jednoduché riadenie, vysoká odolnosť voči nečistotám a nízka cena.

POĎAKOVANIE

Táto práca bola podporovaná Agentúrou na podporu výskumu a vývoja na základe Zmluvy č. APVV-19-0210.

Táto práca vznikla s podporou grantu VEGA 1/0493/19.

LITERATÚRA

- [1] JURIŠICA, L. – HUBINSKÝ, P. – KARDOŠ, J. Robotika. Bratislava : STU, 2005. 134 s.
- [2] H-mostík [online] dostupná na internete < <https://sk.technolong.com/advice/4207-what-is-an-h-bridge-how-it-works-and-how-to-use> >
- [3] Haberkorn [online] < <https://eshop.haberkorn.cz/sk/pohonne-elementy/ozubene-remene/> >
- [4] Rem-technik [online] < <https://www.rem-technik.cz/pohyb-pohony-prevody/prevodovky/pohon-s-ozubenym-hrebenem/system-pridavneho-pastorku-zv-a-hrebenu-844.html> >
- [5] Bibus [online] < <https://www.bibus.sk/prehľad-produktov/mechatronika/gulickove-skrutky-vodiace-skrutky/thomson-metricke-gulickove-skrutky/> >
- [6] CNC-príslušenstvo [online] < <https://www.cncprislusenstvo.sk/ako-na-stavbu-cnc-stroja/linearne-vedenia/> >
- [7] Hennlich [online] < <https://www.hennlich.sk/produkty/motory-a-pohony-linearne-jednotky-s-pohonom-3047/kompaktna-jednotka-so-servomotorom-ty.html> >

Návrh GUI pre ovládanie pohonu

¹Peter GIROVSKÝ, ²Viktor GAMRÁT

^{1,2} Katedra elektrotechniky a mechatroniky, Fakulta elektrotechniky a informatiky,
Technická univerzita v Košiciach

¹peter.girovsky@tuke.sk, ²viktor.gamrat@student.tuke.sk

Abstrakt — Cieľom tohto príspevku je predstaviť návrh riadenia a ovládaním servopohonu Dynamixel cez grafické používateľské rozhranie vytvorené v programe MATLAB. Pre riešenie práce boli použité servopohony Dynamixel AX-12A. V rámci práce bola pre komunikáciu servopohonov a programu MATLAB využívaná sériová komunikácia. V závere príspevku je prezentované navrhnuté GUI pre ovládanie servopohonov.

Kľúčové slová — GUI-grafické užívateľské prostredie, servopohon, sériová komunikácia, MATLAB

I. ÚVOD

Využívanie servopohonov je v rôznych oblastiach priemyslu široko využívané či sa už jedná o robotické aplikácie alebo rôzne mechatronické zariadenia. Pre tieto servopohony existujú riadiace zariadenia, ktoré ponúkajú hotové komplexné riešenia pre ovládanie, riadenie, či sledovanie prevádzkových veličín. Grafické užívateľské rozhranie predstavuje zase silný nástroj na to, aby sa dali nielen servopohony ale aj iné zariadenia jednoducho a intuitívne ovládať. Spojením týchto častí, t.j. servopohonu, riadenia a GUI dostávame komplexné riešenie pre prácu s elektrickými pohonmi. Ako už bolo spomenuté existuje veľa komerčných riešení takéhoto typu, ale pre potreby základného pochopenia princípov práce so servopohonmi sme navrhli jednoduchý systém pre prácu so servopohonmi založený na GUI, ktoré bolo vytvorené v MATLABe.

II. SERVOPOHONY DYNAMIXEL

Dynamixel je vysokovýkonný servopohon určený pre konfigurácie robotov s viacerými kĺbmi vyvinutý kórejskou firmou ROBOTIS. Servopohony Dynamixel sa vyrábajú v širokej škále výkonov a sú použiteľné v rámci rôznych aplikácií od jednoduchých robotov až po zložitejšie robotické celky predstavujúce humanoidných robotov. Každá jednotka Dynamixel je vybavená jednosmerným motorom, regulátorom, ovládačom, senzorom, redukčným prevodom a sieťou integrovanou do jedného modulu. Pre účely tvorby GUI bol zvolený servopohon AX-12A, ktorý ponúka dostatočné možnosti v rámci riešenia danej úlohy. Základné parametre tohto servopohonu sú uvedené v Tabuľke 1.

Tabuľka 1
Parametre servopohonu Dynamixel AX-12A

AX-12A	
Hmotnosť	54.6 g
Rozmery	32 mm x 50 mm x 40 mm
Vstupné napätie	9.0 V ~ 12.0 V (odporúčané: 11.1 V)
Prevodový pomer	254 : 1
Presnosť	0.29°
Rozsah natočenia	0° ~ 300°, nepretržité otáčanie
Zaťažovací moment	1.5 Nm (pri 12 V, 1.5 A)
Rýchlosť naprázdno	59 rpm (pri 12 V)
Prenosová rýchlosť	7843 bps ~ 1 Mbps

Pre komunikáciu servopohonu Dynamixel s počítačom sa používa USB2Dynamixe, cez ktorý je možné prostredníctvom USB portu počítača komunikovať so servopohonom tromi spôsobmi:

- Pomocou TTL
- Pomocou RS485

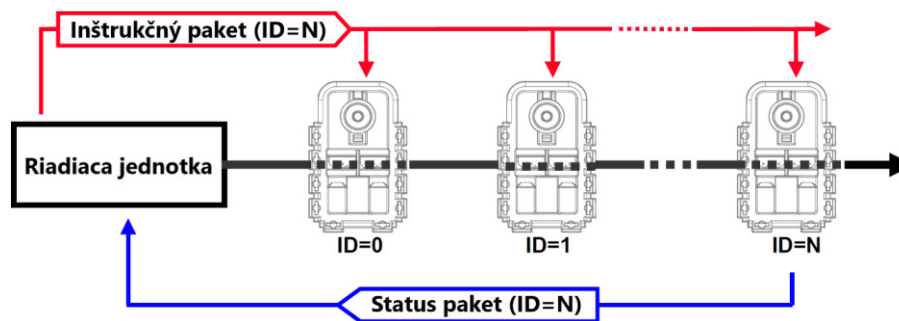
- Pomocou sériovej linky

A. Komunikačný protokol

So servopohonmi Dynamixel je možné komunikovať prostredníctvom inštrukčných paketov:

- Inštrukčný paket - posiela riadiaca jednotka servopohonu Dynamixelu a skladá sa z nasledujúcich prvkov
 - 0xFF, 0xFF – Dva bajty označujúce začiatok prichádzajúceho paketu.
 - ID – Jedinečné identifikačné číslo servopohonu Dynamixel. K dispozícii je 254 dostupných ID hodnôt od rozsahu 0x00 do 0xFD. Adresa 0xFE je Broadcasting ID, ktorá označuje všetky pripojené servopohony Dynamixel.
 - LENGTH – Dĺžka paketu, ktorého hodnota sa vypočíta ako Počet parametrov + 2.
 - INSTRUCTION – Inštrukcie, ktoré má vykonať servopohon.
 - PARAMETER0...N – Používa sa, ak je potrebné poslať ďalšie informácie okrem samotnej inštrukcie.
 - CHECK SUM – Kontrolný súčet, ktorý sa vypočíta ako ID + Dĺžka + Inštrukcia + Parameter1 + ... Parameter N.
- Status paket - posiela servopohon Dynamixel riadiacej jednotke sa skladá z nasledujúcich prvkov
 - 0xFF, 0xFF – Dva bajty označujúce začiatok paketu
 - ID – Jedinečné ID servopohonu Dynamixel, ktorý vracia paket. Počiatočná hodnota je nastavená na 1.
 - LENGTH – Dĺžka paketu, ktorého hodnota je Počet parametrov + 2.
 - ERROR – Bajt reprezentujúci chyby odoslané zo servopohonu. Význam každého bitu je zobrazený v Tabuľke 2.
 - PARAMETER0...N – Používa sa ak je potrebná dodatočná informácia.
 - CHECK SUM – Kontrolný súčet, ktorý sa vypočíta ako ID + Dĺžka + Inštrukcia + Parameter1 + ... ParameterN.

Na Obr. 1 je zobrazená principiálna schéma komunikácie medzi riadiacou jednotkou a servopohonmi Dynamixel.



Obr. 1 Komunikácia s viacerými servopohonmi [12] i

Pre riešenie detekcie kolíznych zostáv boli navrhnuté programy Kamzík a Jastrab, ktoré sú popísané v nasledujúcich častiach článku.

III. GRAFICKÉ ROZHRAŇIE V PROGRAME MATLAB

Pre ovládanie servopohonu AX-12A bolo navrhnuté GUI, ktoré je zobrazené na Obr. 5. Toto rozhranie je rozdelené do troch častí:

- 1 - Nastavovacia časť
- 2 - Ovládacia časť
- 3 - Zobrazovacia časť

A. Nastavovacia časť

Táto časť GUI, zobrazená na Obr. 2, slúži na nastavenie základných parametrov servopohonu:

- Rýchlosť – zadáva sa v otáčkach za minútu v rozsahu od 1 do 100.
- Začiatočná poloha
- Koncová poloha

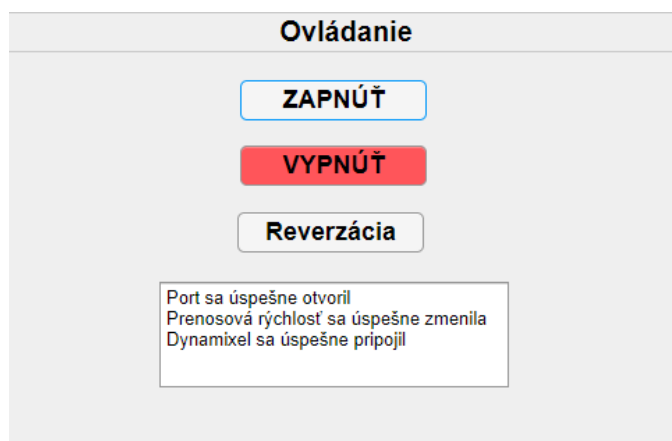
- Obmedzenie momentu – udáva koľko percent z maximálneho momentu servopohon môže použiť pričom rozsah je od 1 do 100.
- Rozpätie rampy – predstavuje chybu medzi cieľovou pozíciou a aktuálnou polohou, zadáva sa v rozsahu od 0 do 255.
- Sklon rampy – nastavuje úroveň momentu v blízkosti cieľovej pozície a nastavuje sa v 7 krokoch, pričom čím vyššia hodnota, tým je možné dosiahnuť väčšiu flexibilitu.
- LED – určuje, či je dióda vypnutá alebo stále zapnutá.
- Režim – vyberá sa z dvoch možností: Joint a Wheel. V prvom prípade sa servopohon otáča medzi dvomi polohami a v tom druhom sa otáča nepretržite v jednom alebo druhom smere.



Obr. 2 Nastavovacia časť GUI

B. Ovládacia časť

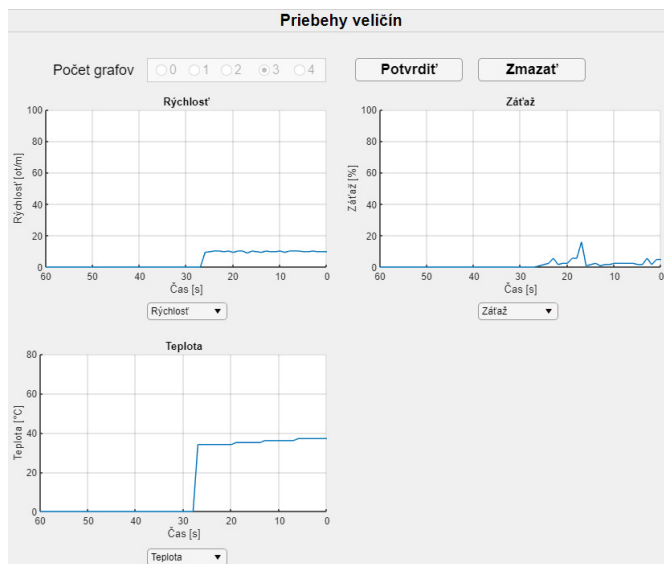
Táto časť obsahuje 3 tlačidlá a jednu textovú oblasť a je zobrazená na Obr.3. Tlačidlo ZAPNÚŤ udáva servopohon do pohybu a obmedzí nastavovať niektoré parametre v predchádzajúcej časti tak, že s jednotlivými objektami používateľ nemôže interagovať. S tlačidlom VYPNÚŤ sa servopohon zastavuje a používateľovi opäť umožní nastavovať všetky parametre. Ak je nastavený režim Wheel tak tlačidlo Reverzácia bude aktívne a používateľ môže meniť smer servopohonu počas jeho chodu. Textová oblasť zobrazuje úspešnosť resp. neúspešnosť otvorenia portu, nastavenia prenosovej rýchlosti a pripojenia servopohonu.



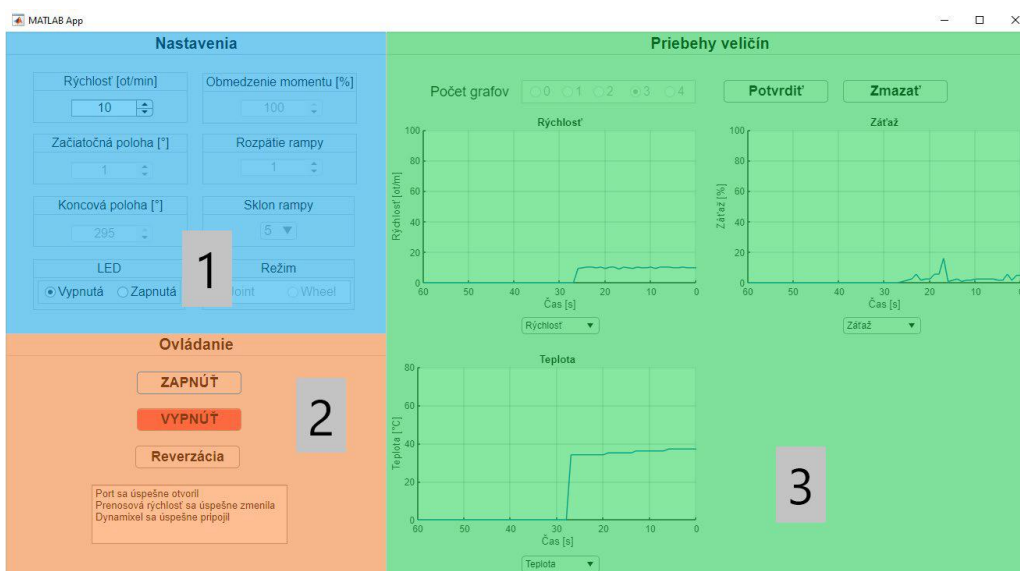
Obr. 3 Ovládacia časť GUI

C. Zobrazovacia časť

Táto časť umožňuje používateľovi vybrať 0-4 priebehy pre sledovanie konkrétnej výstupnej veličiny servopohonu. Zobrazovacia časť GUI je uvedená na Obr.4. Ak je zvolený počet, tlačidlom Potvrdiť sa zobrazia príslušné grafy. Každý graf potom pod sebou bude obsahovať zoznam, ktorý obsahuje výber zo štyroch veličín: rýchlosť, teplota, záťaž a poloha. Ak sú počas chodu servopohonu priebehy zobrazené a následné je pohon zastavený, stlačeným tlačidlom Zmazať sa vymažú dané priebehy.



Obr. 4 Zobrazovacia časť GUI



Obr. 5 GUI pre ovládanie servopohonu

IV. ZÁVER

Cieľom tohto článku bolo predstaviť grafické používateľské rozhranie v programe MATLAB, ktoré umožňuje ovládať servopohon Dynamixel a meniť jeho parametre. Pre komunikáciu servopohonu s programom MATLAB bolo potrebné využiť knižnice, ktoré obsahujú riadiace funkcie na ovládanie servopohonu Dynamixel a nastaviť port pre správnu komunikáciu. Výsledné GUI bolo rozdelené na tri časti – nastavovacia, ovládacia a zobrazovacia. V nastavovacej časti si používateľ môže nastaviť rôzne parametre servopohonu ako je rýchlosť, obmedzenie momentu, nastavenie rampy a pod. Ovládacia časť obsahovala tlačidlá, ktoré zapínali a vypínali chod servopohonu. Implementované bolo aj tlačidlo, ktoré umožňovalo reverzáciu servopohonu. Posledná tretia časť rozhrania zobrazovala priebehy veličín, pričom používateľ si mohol zvoliť koľko grafov chce zobrazit' (najviac štyri) a pri každom grafe si určiť aká veličina sa bude zobrazovať.

POĎAKOVANIE

Táto práca bola podporovaná Agentúrou na podporu výskumu a vývoja na základe Zmluvy č. APVV-19-0210.

Táto práca vznikla s podporou grantu VEGA 1/0493/19.

LITERATÚRA

- [1] Graphical user interface, [online]. Dostupné na internete: <https://www.britannica.com/technology/graphical-user-interface>
- [2] ZAPLATÍLEK, Karel – DOŇAR, Bohuslav: MATLAB tvorba uživatelských aplikací. 1. vyd. Praha: BEN, 2004. ISBN 80-7300-133-0
- [3] Využití grafického uživatelského rozhraní MATLABU ve výzkumu a výuce měření, [online]. Dostupné na internete: http://dsp.vscht.cz/konference_matlab/matlab03/blaska.pdf
- [4] MATLAB GUI, [online]. Dostupné na internete: <https://www.mathworks.com/discovery/matlab-gui.html>
- [5] MATLAB App Building, [online]. Dostupné na internete: https://www.mathworks.com/help/pdf_doc/matlab/creating_guis.pdf
- [6] ROBOTIS Dynamixel, [online]. Dostupné na internete: <https://www.robotis.us/dynamixel/>
- [7] Tribotix Dynamixels, [online]. Dostupné na internete: <https://tribotix.com/dynamixels/>
- [8] Dynamixel AX-12A e-Manual, [online]. Dostupné na internete: <https://emanual.robotis.com/docs/en/dxl/ax/ax-12a/>
- [9] USB2DYNAMIXEL e-Manual, [online]. Dostupné na internete: <https://emanual.robotis.com/docs/en/parts/interface/usb2dynamixel/>
- [10] USB2DYNAMIXEL User's manual, [online]. Dostupné na internete: <https://www.pishrobot.com/files/products/datasheets/usb2dynamixel.pdf>
- [11] DYNAMIXEL Wizard 2.0 e-Manual, [online]. Dostupné na internete: https://emanual.robotis.com/docs/en/software/dynamixel/dynamixel_wizard2/
- [12] Dynamixel AX-12 manual, [online]. Dostupné na internete: <https://www.manualslib.com/manual/13679/Robotis-Dynamixel-Ax-12.html#manual>
- [13] Matlab PortHandler, [online]. Dostupné na internete: https://emanual.robotis.com/docs/en/software/dynamixel/dynamixel_sdk/api_reference/matlab/matlab_porthandler/#matlab-porthandler
- [14] Matlab PacketHandler, [online]. Dostupné na internete: https://emanual.robotis.com/docs/en/software/dynamixel/dynamixel_sdk/api_reference/matlab/matlab_packethandler/#matlab-packethandler

Analýza vplyvu zmeny dimenzionality na efektívitu reprezentácie geometrie voxelizovaných 3D scén

¹Matej Kormanský, ²Branislav Madoš

^{1,2}Katedra počítačov a informatiky, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

¹matej.kormansky@student.tuke.sk, ²branislav.mados@tuke.sk,

Abstrakt – Táto práca študuje efekty škálovania dimenzionality na hierarchické údajové štruktúry sparse voxel octrees a sparse voxel directed acyclic graphs, ktoré sa využívajú na reprezentáciu geometrie trojrozmerných voxelizovaných scén. V rámci práce bol navrhnutý a implementovaný algoritmus na konštrukciu sparse voxel octrees štruktúry, ktorý vychádza z Baertovho algoritmu a obohacuje ho o škálovanie dimenzionality a program na konštrukciu sparse voxel directed acyclic graphs štruktúry zlúčením rovnakých podstromov štruktúry sparse voxel octrees. Funkčnosť algoritmov a efekty škálovania dimenzionality na tieto hierarchické dátové štruktúry boli otestované na rôznych modeloch voxelizovaných do rôznych rozlíšení. Práca porovnáva efekty škálovania dimenzionality na kompresný pomer a rýchlosť traverzovania týchto údajových štruktúr.

Kľúčové slová – hierarchické údajové štruktúry, bezstratová kompresia, škálovanie dimenzionality, SVO, SVDAG

I. ÚVOD

V počítačovej grafike sa na reprezentáciu geometrie scény používa trojdimenzionálna sieť skalárnych hodnôt s veľkosťou 1 bit na jeden voxel, kde pasívny voxel je reprezentovaný ako hodnota bitu 0 a aktívny voxel ako hodnota bitu 1. Takáto reprezentácia geometrie trojrozmerných scén má však svoju nevýhodu. Pri vysokom rozlíšení scény môže veľkosť trojrozmernej súradnicovej siete voxelov byť v rádoch gigabajtov, ba až terabajtov. Preto je potrebné tieto dáta skomprimovať tak, aby sa dosiahla nižšia náročnosť na ukladací priestor. Populárnym riešením tohto problému sú hierarchické dátové štruktúry, ktoré umožňujú bezstratovú kompresiu reprezentácie geometrie scény.

Jednou z nich je oktantový strom Sparse Voxel Octrees (SVO) [1][2], ktorej prínos spočíva v eliminácii prázdnych podoblastí scény, čo umožňuje jej bezstratovú kompresiu. Ďalšou štruktúrou je orientovaný acyklický graf Sparse Voxel Directed Acyclic Graphs (SVDAG) [3], ktorý umožňuje dosiahnuť dodatočnú kompresiu oproti SVO cez zlučovanie rovnakých podstromov resp. podgrafov a orientovaný acyklický graf Symmetry-aware Sparse Voxel Directed Acyclic Graphs (SSVDAG) [4], ktorý dosahuje dodatočnú kompresiu zlučovaním symetrických podgrafov na základe transformácií zrkadlením. Hoci SVO a SVDAG poskytli slušný kompresný pomer a efektívne traverzovanie ich štruktúr, stále je tu iniciatíva vylepšovať ich vlastnosti. Napríklad boli vytvorené štruktúry Pointerless Sparse Voxel Octrees (PSVO) a Pointerless Sparse Voxel Directed Acyclic Graphs (PSVDAG) [5], ktoré umožňujú konštrukciu oktantového stromu a orientovaného acyklického grafu bez použitia smerníkov, čo ďalej vylepšilo kompresný pomer, ale urobilo prehľadávanie týchto štruktúr ťažším.

V roku 2021 bola skúmaná metóda škálovania dimenzionality [6], ktorá má potenciál ďalej vylepšiť kompresiu a vlastnosti hierarchických dátových štruktúr (HDS) SVO a SVDAG. Efekty škálovania dimenzionality na reprezentáciu voxelizovaných 3D scén však neboli ešte podrobne preskúmané.

II. ŠKÁLOVANIE DIMENZIONALITY

Zmysel škálovania dimenzionality spočíva v tom, že ku voxelom, ktoré popisujú jednotlivé body scény pre trojrozmerný priestor sa pristupuje akoby popisovali body nie trojrozmerného priestoru, ale priestoru s vyšším počtom rozmerov. To môže ovplyvniť výslednú formu hierarchickej štruktúry a jej vlastnosti. Škálovanie dimenzionality dokáže ovplyvniť počet uzlov, veľkosť masky dcérskych uzlov, počet smerníkov, hĺbku stromu resp. orientovaného acyklického grafu.

Príkladom môže byť dvojrozmerná scéna s veľkosťou 4×4 pixelov, ktorej body boli linearizované pomocou Mortonovej priestor vyplňajúcej krivky do reťaze Mortonových kódov, kde reprezentované sú iba aktívne voxely (Fig. 1).

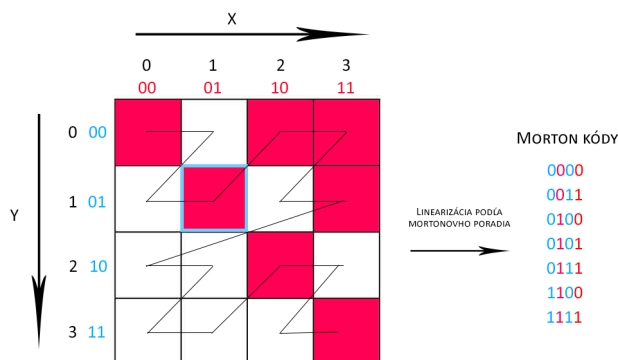


Fig. 1 Linearizácia dvojrozmernej scény pixelov podľa Mortonovej priestor vyplňajúcej krivky.

Fig. 1 popisuje proces linearizácie bodov dvojrozmernej scény pixelov podľa Mortonovho poradia. Pixely vyfarbené červenou farbou sú aktívne a pixely vyfarbené bielou farbou sú pasívne. Mortonov kód bodu sa vypočíta prekladaním bitov reprezentácie jeho súradníc. Napríklad hodnota Mortonovho kódu aktívneho pixelu nachádzajúceho sa na súradniciach $x = 01_2 = 1_{10}$ a $y = 01_2 = 1_{10}$ je hodnota $m = 0011_2 = 3_{10}$. Bity reprezentujúce hodnoty súradníc bodu sú striedavo použité tak, že po bite súradnice x (označený červenou farbou) nasleduje bit súradnice y (označený modrou farbou) v tomto prípade v smere sprava doľava od LSB po MSB.

Mortonov kód umožňuje podľa jeho bitovej reprezentácie odvodiť v ktorej podoblasti scény (oktante v 3D, kvadrante v 2D) sa bod nachádza. Zoberme do úvahy príklad dvojrozmernej scény pixelov a bod so súradnicami $x = 1$, $y = 1$, označený modrým orámovaním na Fig. 1. Bity Mortonovho kódu tohto bodu usporiadame do dvojíc. Prvý bit z tejto dvojice je bit súradnice x a druhý bit je bit súradnice y (od LSB po MSB) (Fig. 2).

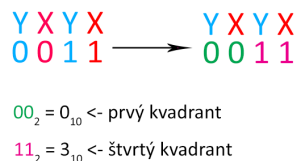


Fig. 2 Usporiadanie bitov Mortonovho kódu do dvojíc pre 2D.

Takto usporiadané dvojice bitov Mortonovho kódu označujú kvadranty a subkvadranty podľa Mortonovho poradia, v ktorých sa nachádza bod $x = 1$, $y = 1$. Dvojica bitov 00 označuje prvý kvadrant dvojrozmernej scény, zatiaľ čo dvojica bitov 11 označuje štvrtý subkvadrant prvého kvadrantu dvojrozmernej scény pixelov, ktorým je bod $x = 1$, $y = 1$, podľa Mortonovho poradia, tak ako je vyobrazené na Fig. 3.

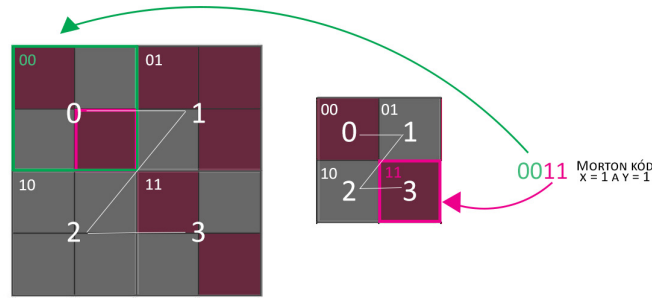


Fig. 3 Kvadranty dvojrozmernej scény, na ktoré referujú dvojice bitov Mortonovho kódu bodu $x = 1$, $y = 1$.

Z bitovej reprezentácie Mortonovho kódu sa dá zistiť, do ktorého oktantu (kvadrantu v 2D) bod patrí. Na základe Mortonovych kódov môžeme pristúpiť k budovaniu hierarchickej údajovej štruktúry. Pre jednoduchosť použijeme príklad dvojrozmernej scény pixelov a Mortonove kódy z príkladu na Fig. 1 na transformáciu do podoby SVO údajovej štruktúry (v 2D kvadrantové stromy). Na základe dvojíc bitov obsiahnutých v Mortonovych kódoch aktívnych pixelov dvojrozmernej scény vieme určiť, ktoré bity masky dcérskych uzlov majú mať hodnoty 1 alebo 0, keďže bity masky označujú kvadranty, ktoré obsahujú aktívne pixely, stačí ak nastavíme jednotkové bity v maske na pozíciách podľa hodnoty dvojíc bitov v Mortonovych kódoch aktívnych pixelov. Následne po prepojení rodičovských uzlov s dcérskymi uzlami prostredníctvom smerníkov získame SVO štruktúru, ktorá predstavuje dvojrozmernú scénu pixelov (Fig. 4).

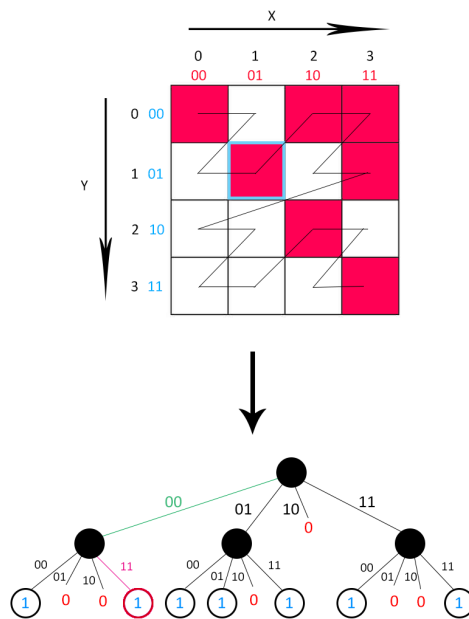


Fig. 4 Grafická podoba SVO stromu reprezentujúca 2D scénu z príkladu na Fig. 1.

Pri škálovaní dimenzionality scény pristupujeme k bodom scény ako k bodom, ktoré patria do priestoru s rozdielnym počtom rozmerov v porovnaní s počtom rozmerov pôvodnej scény. V príklade na Fig. 4 sme zostrojili hierarchickú údajovú štruktúru kvadrantový strom, ktorý kóduje dvojrozmernú scénu pixelov na základe ich Mortonovych kódov. Predpokladajme, že tieto body dvojrozmernej scény patria do trojrozmerného priestoru. To znamená, že každý Mortonov kód je zostrojený prekladaním bitov nie dvoch súradníc, ale troch súradníc aj napriek tomu, že tretia súradnica prakticky neexistuje. Tak ako sme bity Mortonovho kódu pri 2D usporiadávali do dvojíc bitov, kde prvý bit z tejto dvojice je bit súradnice x a druhý bit je bit súradnice y , tak pri 3D budeme tieto bity usporiadať do trojíc, kde prvý bit označíme ako bit súradnice x , druhý bit ako bit súradnice y a tretí bit ako bit tretej súradnice z (od LSB po MSB). Príklad pre Morton kód bodu $x = 2$, $y = 2$ sa nachádza na Fig. 5.

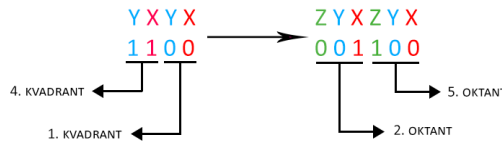


Fig. 5 Porovnanie dvojíc a trojíc bitov Mortonovho kódu bodu $x = 2, y = 2$ pri 2D a 3D škálovaní.

Pri 2D sa na určenie pozície bodu v kvadrante používali dva bity, takže existovalo 2^2 rôznych kvadrantov. Pri 3D je pozícia bodu zakódovaná na troch bitoch, čo dokopy dáva 2^3 rôznych oktantov. Nech D predstavuje počet rozmerov, potom počet podoblastí N , na ktoré je možné rozdeliť oblasť scény je daný výrazom $N = 2^D$. Tento výraz definuje aj minimálnu veľkosť masky dcérskych uzlov potrebnú na zakódovanie všetkých možných podoblastí scény (kvadranty v 2D, oktanty v 3D atď.). Všimnime si, že zoskupovanie bitov do trojíc v príklade na Fig. 5 zmenilo pozíciu bodu $x = 2, y = 2$ zo štvrtého kvadrantu na druhý oktant a prvého subkvadrantu na piaty suboktant. Už teraz môžeme povedať, že zmena dimenzionality bude viesť k inej podobe HDS aj napriek tomu, že budú reprezentovať rovnakú scénu.

Proces konštrukcie SVO štruktúry z dvojrozmernej scény pixelov, ktorý bol vykonaný vyššie (Fig. 4) zopakujeme s tým, že aplikujeme 3D škálovanie dimenzionality, kde budeme zoskupovať bity Mortonovych kódov do trojíc, ktoré budú diktovať pozíciu bodu v maske dcérskych uzlov. Po dokončení procesu konštrukcie SVO s 3D škálovaním dimenzionality získavame rozdielnu HDS ako tomu bolo pri 2D. Výsledná SVO štruktúra pri aplikovaní 3D škálovania dimenzionality sa nachádza na Fig. 6.

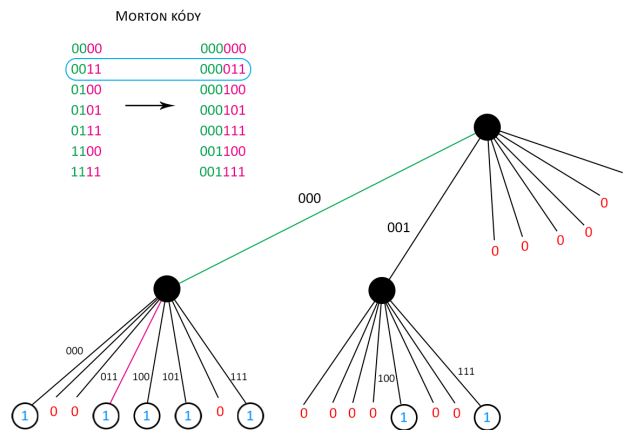


Fig. 6 SVO strom reprezentujúci 2D scénu z príkladu na Fig. 1 s aplikáciou škálovania dimenzionality do 3D.

Škálovanie dimenzionality umožnilo v tomto prípade zredukovať počet uzlov zo štyroch na tri a počet smerníkov z troch na dva za cenu väčšej veľkosti uzla, keďže maska dcérskych uzlov musí mať veľkosť osem bitov. Ak predpokladáme, že masky sú zarovnané na 32 bitov a veľkosť smerníkov je taktiež 32 bitov, tak veľkosť pôvodnej štruktúry v bitoch je $4 * 32 + 3 * 32 = 224b$ (4 uzly, 3 smerníky) a veľkosť štruktúry pri aplikácii 3D škálovania dimenzionality je $3 * 32 + 2 * 32 = 160b$ (4 uzly, 3 smerníky). V tomto prípade sme zredukovali veľkosť výslednej štruktúry o 64 bitov. V prípade, že masky nie sú zarovnané na 32 bitov, čiže majú veľkosť štyri bity resp. osem bitov je veľkosť pôvodnej štruktúry $4 * 4 + 3 * 32 = 112b$ a veľkosť štruktúry pri aplikácii 3D škálovania dimenzionality $3 * 8 + 2 * 32 = 88b$, čo predstavuje redukcii veľkosti pôvodnej štruktúry o 24 bitov. Cieľom škálovania dimenzionality je dosiahnutie kompromisu medzi znižovaním počtu uzlov a smerníkov a zvyšovaním veľkosti uzlov, tak aby sa dosiahol vyšší kompresný pomer a neustratila sa schopnosť efektívne vyhľadávať údaje.

III. ALGORITMUS PRE KONŠTRUKCIU SVO

V roku 2013 Baert et al. publikovali prácu [7], kde popísali proces voxelizácie modelu z trojuholníkovej polygónovej siete do trojrozmernej voxelizovanej siete a následne proces konštrukcie SVO z reprezentácie voxelizovanej scény v Mortonovom poradí. Výsledkom bol out-of-core algoritmus na konštrukciu SVO údajovej štruktúry z reprezentácie voxelizovanej scény v Mortonovom kóde, ktorý umožnil transformovať modely vysokého rozlíšenia pri spotrebe nízkeho množstva

pamäte do SVO údajovej štruktúry [7]. Tento algoritmus však nepodporuje škálovanie dimenzionality a z Mortonových kódov produkuje oktantový strom. Tento algoritmus je veľmi efektívny a vďaka svojej out-of-core povahe dokáže spracovať linearizované scény voxelov v Mortonovom poradí veľkých rozmerov. Preto náš algoritmus vychádza z Beartovho algoritmu pre konštrukciu SVO z Mortonových kódov a obohacuje ho o funkciu škálovania dimenzionality.

Počet dimenzií, na ktorý sa bude hierarchická štruktúra škálovať bude vstupným parametrom programu. Vstupný parameter počtu rozmerov je kladné celé číslo väčšie ako dva. Počet rozmerov rovný trom produkuje klasický oktantový strom ako pôvodný Baertov algoritmus. Vstupný súbor Mortonových kódov pozostáva zo sekvencie 64-bitových celých čísel bez znamienka usporiadaných od najmenšieho po najväčšie. Výstupom programu je hierarchická štruktúra SVO so škálovaním dimenzionality, ktorá je uložená v súboroch s príponou *.svo. Každý súbor obsahuje uzly, ktoré sa nachádzajú na rovnakej vrstve stromu.

IV. ALGORITMUS PRE KONŠTRUKCIU SVDAG

Na konštrukciu SVDAG sme použili metódu zdola nahor (bottom-up algoritmus), ktorá postupuje od listových uzlov smerom ku koreňovému uzlu. Metóda zdola nahor sa skladá z nasledujúcich krokov [3]

- Najprv sa zlúčia identické listové uzly, ktoré majú rovnakú masku dcérskych uzlov.
- Smerníky na dcérske uzly na vrstve vyššie budú aktualizované tak, aby referovali na nové jedinečné listové uzly.
- Potom sa na vyššej vrstve nájdu všetky uzly, ktoré zdieľajú hodnoty masky dcérskych uzlov a smerníkov na dcérske uzly. Takéto uzly sú koreňovými uzlami identických podstromov a môžu byť zlúčené.
- Následne je potrebné znova upraviť hodnoty smerníkov na nadradenej vrstve, aby správne referovali na koreňové uzly jedinečných podstromov.
- Takto sa iteratívne postupuje po vrstvách smerom ku koreňu stromu, až kým sa nenájdu všetky identické podstromy a nezlúčia sa do jedného unikátneho podstromu.

Vstupnými parametrami algoritmu sú súbory SVO hierarchickej štruktúry s príponou *.svo, počet rozmerov, na ktorý sa HDS bude škálovať a cesta k adresáru do ktorého sa budú zapisovať súbory výslednej HDS. Hodnota vstupného parametra počtu rozmerov musí byť rovnaká ako dimenzionalita vstupnej SVO hierarchickej štruktúry. Výstupom programu je hierarchická štruktúra SVDAG so škálovaním dimenzionality, ktorá je uložená v súboroch s príponou *.dag. Každý súbor obsahuje uzly, ktoré sa nachádzajú na rovnakej vrstve acyklického grafu.

V. VYHODNOTENIE

Vyhodnotenie a overenie správnosti implementácie algoritmov pre zostavovanie SVO a SVDAG hierarchických štruktúr so škálovaním dimenzionality sme vykonali na siedmich modeloch scén voxelov s rôznym rozlíšením. Výpočty a konštrukcia údajových štruktúr SVO a SVDAG boli realizované na systéme s procesorom Intel Core i7-7700HQ 2.80GHz, 16 GB RAM, 128 GB SSD a GPU Nvidia GeForce GTX 1050 Ti.

Fig. 7 zobrazuje priemerný kompresný pomer, ktorý dosiahli štruktúry SVO a SVDAG pri rôznych hodnotách dimenzionality.

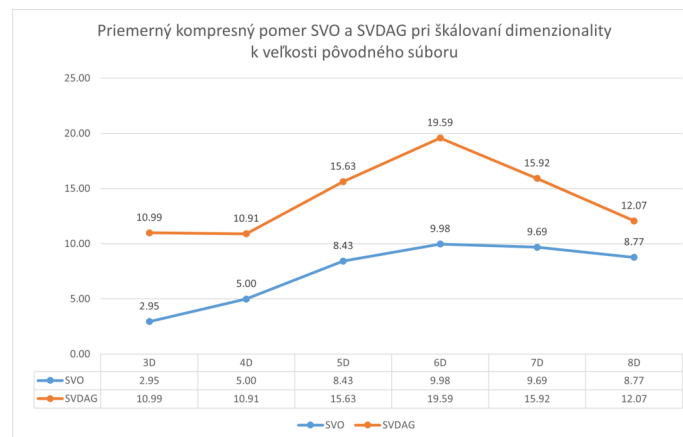


Fig. 7 Priemerný kompresný pomer SVO a SVDAG pri škálovaní dimenzionality k veľkosti pôvodného súboru.

Najlepší kompromis medzi redukciou počtu uzlov, smerníkov a zvyšovaním veľkosti uzlov bol dosiahnutý pri 6D škálovaní dimenzionality, pri ktorom dosiahli údajové štruktúry SVO a SVDAG svoju najnižšiu veľkosť, a teda najvyšší kompresný pomer 9.98 pri SVO a 19.59 pri SVDAG. Pri vyšších hodnotách dimenzionality ako 6D začíname pozorovať nárast veľkosti údajových štruktúr.

Fig. 8 vyobrazuje priemerné hodnoty zrýchlenia traverzovania SVO a SVDAG pri rôznych hodnotách dimenzionality v porovnaní s klasickou 3D verziou SVO a SVDAG.

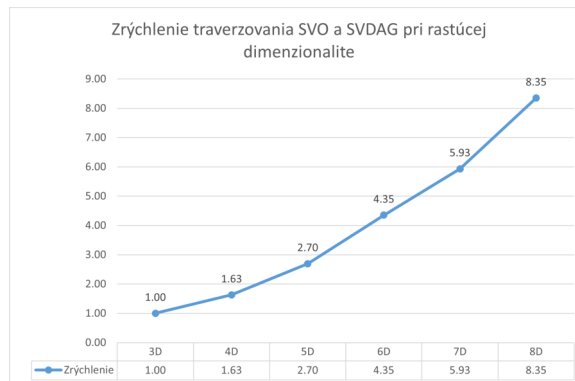


Fig. 8 Priemerné zrýchlenie traverzovania SVO a SVDAG pri rôznych hodnotách dimenzionality.

Pri rastúcej hodnote dimenzionality sme konzistentne dosahovali vyššie zrýchlenie pri traverzovaní údajových štruktúr SVO a SVDAG. Toto zrýchlenie bolo dôsledkom nižšieho počtu úrovní a uzlov, z ktorých pozostávali štruktúry SVO a SVDAG, a teda nižšieho počtu iterácií traverzovacieho algoritmu.

VI. ZÁVER

Práca prezentuje efekty škálovania dimenzionality na údajové štruktúry SVO a SVDAG. Výsledky testov ukazujú, že pri zvyšovaní dimenzionality dochádza k výraznému poklesu počtu uzlov a adries (smerníkov) údajovej štruktúry SVO. Po vytvorení SVDAG z SVO štruktúry je hierarchická štruktúra dodatočne zbavená redundantných uzlov totožných podstromov, čo ešte zvyšuje kompresný pomer oproti SVO a ďalej znižuje počet uzlov a smerníkov. Pri škálovaní dimenzionality dochádza k redukcii počtu uzlov a adries za cenu väčšej veľkosti uzlov. Preto je potrebné nájsť kompromis medzi redukciou počtu uzlov a zvyšovaním veľkosti uzlov. Najlepšie výsledky boli dosiahnuté pri 6D škálovaní dimenzionality, ktoré dosiahlo v priemere 3.38-krát lepší kompresný pomer ako pri klasickom 3D SVO a 1.78-krát lepší kompresný pomer ako pri klasickom 3D SVDAG. Optimálnou voľbou je 6D škálovanie SVO a SVDAG, ktoré dosahuje v priemere najlepší kompresný pomer.

POĎAKOVANIE

This publication has been published with the support of the Operational Program Integrated Infrastructure within project: Research in the SANET Network and Possibilities of Its Further Use and Development (ITMS code: 313011W988), co-financed by the ERDF. This research was supported by KEGA 002TUKE-4/2021 Implementation of Modern Methods and Education Forms in the Area of Cybersecurity Towards Requirements of Labour Market.

REFERENCES

- [1] S. Laine and T. Karras, "Efficient sparse voxel octrees," in *Proceedings of the 2010 ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games*, ser. I3D '10. New York, NY, USA: Association for Computing Machinery, 2010, pp. 55–63. [Online]. Available: <https://doi.org/10.1145/1730804.1730814>
- [2] B. Madoš, E. Chovancová, and M. Hasin, "Evaluation of pointerless sparse voxel octrees encoding schemes using Huffman encoding for dense volume datasets storage," in *2020 18th International Conference on Emerging eLearning Technologies and Applications (ICETA)*, 2020, pp. 424–430.
- [3] V. Kämpe, E. Sintorn, and U. Assarsson, "High resolution sparse voxel dags," *ACM Trans. Graph.*, vol. 32, no. 4, 7 2013. [Online]. Available: <https://doi.org/10.1145/2461912.2462024>
- [4] A. J. Villanueva, F. Marton, and E. Gobbetti, "Ssvdags: Symmetry-aware sparse voxel dags," in *Proceedings of the 20th ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games*, ser. I3D '16. New York, NY, USA: Association for Computing Machinery, 2016, pp. 7–14. [Online]. Available: <https://doi.org/10.1145/2856400.2856420>
- [5] L. Vokorokos, B. Madoš, and Z. Bilanová, "Psvdag: Compact voxelized representation of 3d scenes using pointerless sparse voxel directed acyclic graphs," *Computing and Informatics*, vol. 39, pp. 587–616, 2020.
- [6] H. Khorshidiyeh, B. Madoš, and A. BALÁŽ, "Kompresia multi-dimenzionálnych dát s využitím transformácie ich dimenzionality," in *Electrical Engineering and Informatics 12: Proceedings of the Faculty of Electrical Engineering and Informatics of the Technical University of Košice*. Technická univerzita v Košiciach, 2021, pp. 165–170.
- [7] J. Baert, A. Lagae, and P. Dutré, "Out-of-core construction of sparse voxel octrees," in *HPG '13*, 2013.

Návrh mikropásikovej antény pre použitie v hybridnom FSO/RF systéme

¹Maroš Lapčák, ²Luboš Ovseník, ³Jakub Oravec, ⁴Norbert Zdravecký, ⁵Samuel ANDREJČÍK

^{1,2,3,4,5} Katedra elektroniky a multimediálnych telekomunikácií, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

¹maros.lapcak@tuke.sk, ²lubos.ovsenik@tuke.sk, ³jakub.oravec@tuke.sk,
⁴norbert.zdravecky@tuke.sk, ⁵samuel.andrejcek@tuke.sk

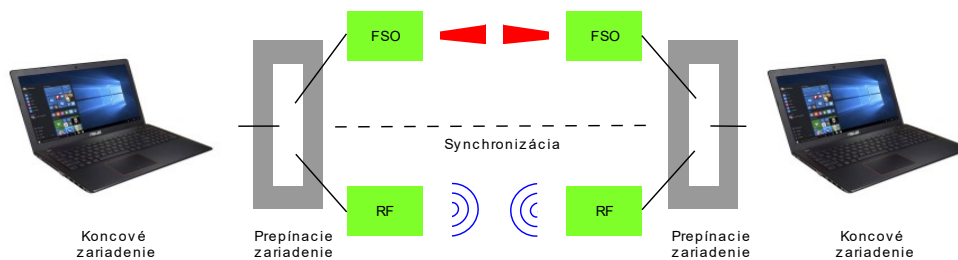
Abstrakt — Táto publikácia sa zaoberá návrhom záložnej RF linky pre hybridný FSO/RF systém. Tento systém pozostáva z primárnej FSO linky a sekundárnej RF linky. Obe komunikačné linky majú rozdielne vplyvy prostredia na ich funkčnosť. V prípade FSO ide najmä o hmlu a v prípade RF linky o dážď a vzájomné rušenie zo strany užívateľov. Preto je potrebné správne navrhnuť a nakonfigurovať záložnú RF linku na prevádzku v nelicencovanom pásme a splniť všetky podmienky pre správnu funkčnosť.

Kľúčové slová — FSO, hybridný FSO/RF systém, mikropásiková anténa, RF komunikácia

I. ÚVOD

FSO systémy sú založené na priamej viditeľnosti, používajú laserové svetlo na zabezpečenie optického spojenia. V súčasnej dobe je FSO systém schopný preniesť až 2,5 Gb/s dátovej, hlasovej a obrazovej komunikácie vzduchom. Táto možnosť zabezpečuje optické pripojenie bez použitia optického kábla alebo prenosu v licencovaných frekvenčných pásmach. FSO systémy potrebujú na prenos svetlo použitím svetelných diód (LED) alebo laserov (zosilnenie svetla stimulovanou emisiou žiarenia). Použitie laserov je podobné ako použitie optického prenosu pomocou káblov z optických vlákien, rozdielom je však médium.

FSO klasifikujeme ako optickú komunikáciu rýchlosťou svetla, keďže sa šíri vzduchom a prenos svetlom je rýchlejší ako prenos sklom. Technológia FSO je pomerne jednoduchá. Pracuje na vlnových dĺžkach od 780 nm do 1600 nm. Na konvertovanie optického signálu na elektrický využívajú optickoelektrické (O/E) konvertory. FSO systémy sa skladajú z troch hlavných komponentov: vysielač, prenosový kanál a prijímač. Úlohou vysielača je vysielať optické žiarenie do atmosféry. Prenosový kanál je tvorený voľným prostredím, v ktorom existujú turbulentné víry (dážď, dym, mraky, hmla, plyny, teplotné rozdiely a aerosól). Prijímač má za úlohu spracovanie prijatého signálu. Dĺžka FSO linky sa pohybuje od 300 m až 5 km, je možné dosiahnuť aj väčšie vzdialenosti. Maximálna dĺžka linky sa pohybuje okolo 16 km, avšak je ovplyvnená rýchlosťou a požadovanou dostupnosťou [1].



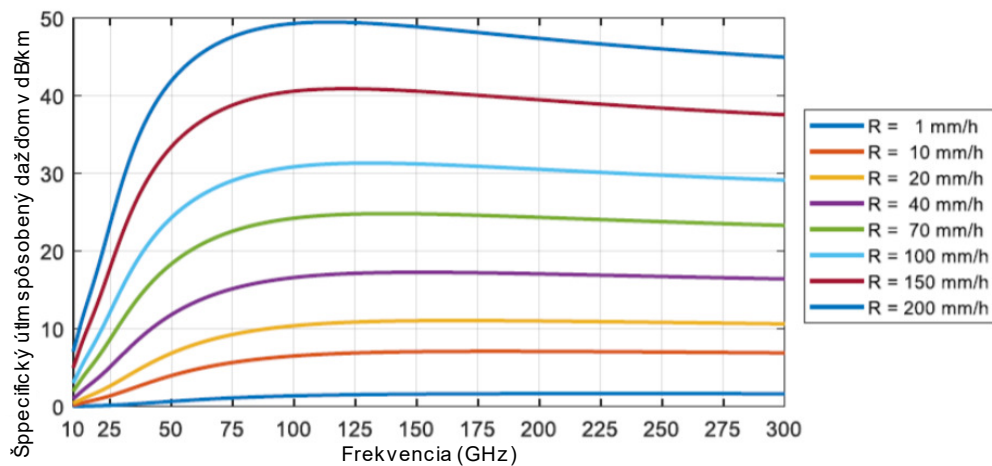
Obr. 1 Příklad fungovania hybridného FSO/RF systému.

Ak uvažujeme o návrhu sekundárnej RF linky pre potreby hybridného FSO/RF systému, je potrebné brať do úvahy niekoľko faktorov, ako sú rozmery antény, vplyvy prostredia na RF komunikáciu, dobrá vysielačská charakteristika antény a tak isto aj cenové náklady na jej výrobu. Ak hovoríme o poveternostných vplyvoch na RF komunikáciu, tak je to predovšetkým:

- atmosférické plynné straty spôsobené predovšetkým kyslíkom a vodnou parou,

- útlm zrážok v dôsledku dažďa,
- straty rozptylom, ktoré zahŕňajú difúzny odraz a zrkadlový odraz,
- difrakcia alebo strata ohybom [2].

Ako je možné vidieť na Obr. 2, útlm spôsobený dažďom je pre každú frekvenciu rozdielny. Preto je potrebné tento atmosférický vplyv dôkladne zväziť pri návrhu každého druhu antény.



Obr. 2 Útlm RF systému vzhľadom na úhm zrážok.

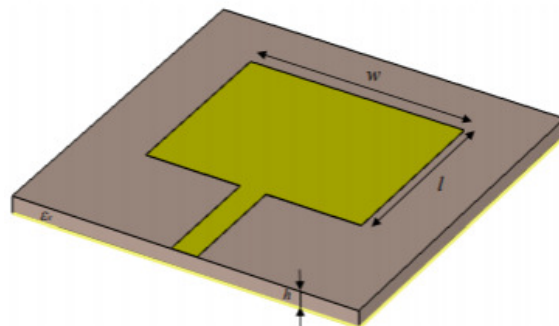
II. MIKROPÁSIKOVÁ ANTÉNA

Mikropásovité alebo patch antény sú čoraz užitočnejšie, pretože je možné ich tlačiť priamo na dosku s plošnými spojmi. Mikropásovité antény sa na trhu s mobilnými telefónmi stávajú veľmi rozšírenými. Patch antény sú lacné, majú nízky profil a sú ľahko vyrobené. Mikropásovité antény sú úzkopásmové. Je vyrobená leptaním anténneho prvku do kovovej stopy viazanej na izolačné dielektrikum substrátu, ako je doska plošných spojov, s plynulou vrstvou kovu viazanou na opačnú stranu podkladu, ktorú tvorí uzemňovacia plocha. Je možné vytvoriť akýkoľvek nepretušovaný tvar. Najčastejšie používané sú však štvorcové, obdĺžnikové, kruhové a eliptické tvary. Vďaka nízkemu profilu sú tieto antény mechanicky odolné a môžu byť upravené pre akékoľvek zakrivenie. Niektoré pásovité antény nepoužívajú dielektrický substrát a preto je výsledná štruktúra menej robustná, ale má väčšiu šírku pásma [3, 4].

A. Typy mikropásových antén

1) Mikropásovité patch antény

Majú rôzne tvary plátka, najčastejšie používanými sú pravouhlé. Zisk ktorý je typický pre tento typ antén sa pohybuje od 5 do 6 dB. Patch antény vyžarujú len do polpriestoru to znamená, že vodivá základňa sa nachádza pod plátkami. Spodná strana substrátu obsahuje zemnú plochu. Dĺžka vodivej plochy w a šírka vodivej plochy l majú dopad na rezonančnú frekvenciu, vstupnú impedanciu a taktiež šírku pásma. Výška h definuje hrúbku dielektrického substrátu. Typická mikropásovité pravouhlá patch anténa je zobrazená na Obr. 3.



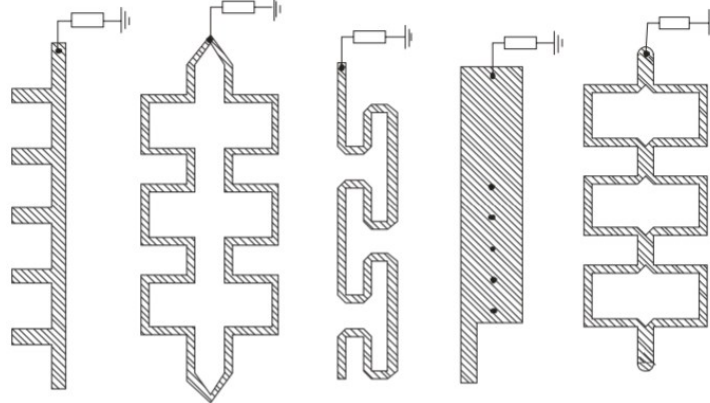
Obr. 3 Mikropásovité pravouhlá patch anténa.

2) Mikropásovité dipól

Vyznačuje sa všesmerovou vyžarovacou charakteristikou a jednoduchou štruktúrou. Šírka dipólu w je považovaná za nulovú. Vyžarovací odpor, šírka pásma, výkonnosť a úroveň krížovej polarizácie sa líši od patch antény. Rovnako majú iba vyžarovacie charakteristiky pri pozdĺžnom rozložení prúdu. Vynikajú svojou veľkosťou a lineárnou polarizáciou [5].

3) Mikropásikové antény s postupujúcou vlnou

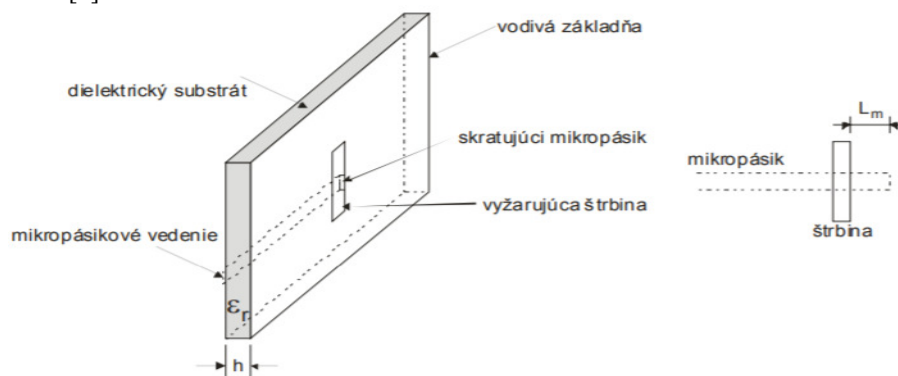
Skladajú sa z reťazených pravidelných úsekov vedenia, alebo z dlhých mikropásikových úsekov s dostatočnou šírkou pásika. Koniec týchto antén je zakončený prispôbenu odporovou záťažou zabraňujúcou vzniku stojatých vln na anténe. Na Obr. 4 sú zobrazené typy periodických štruktúr mikropásikových antén s postupujúcou vlnou [5].



Obr. 4 Typy periodických štruktúr mikropásikových antén s postupujúcou vlnou [4].

4) Mikropásikové štrbinové antény

Mikropásikové štrbinové antény sú zložené zo štrbiny v podkladovej doske uzemneného substrátu, štrbina môže mať rôzne tvary. Najčastejšie používanými tvarmi sú: kruhová štrbina, obdĺžniková štrbina, prstencová štrbina a kužeľová štrbina. Mikropásiková štrbinová anténa je zobrazená na Obr. 5. Napájanie prebieha buď koplanárnym - vlnovodovým vedením, mikropásikovým vedením alebo koaxiálnym káblom. Vyžarujú do celého smeru a teda sú všesmerové [5].



Obr. 5 Mikropásiková štrbinová anténa [4].

B. Výhody a nevýhody mikropásikových antén a spôsoby napájania týchto antén

Mikropásikové antény majú niekoľko výhod v porovnaní s tradičnými mikrovlnnými anténami a preto veľa ich aplikácií pokrýva široké frekvenčné pásmo radovo od 100 MHz do 100 GHz. Výhody mikropásikových antén v porovnaní s tradičnými mikrovlnnými anténami sú:

- Nízka hmotnosť, malé rozmery, nízko profilové.
- Nízke výrobné náklady, preto je možné ich vyrábať vo veľkom množstve.
- Podporujú lineárnu aj kruhovú polarizáciu.
- Schopné vykonávať operácie s dvojitou a trojitou frekvenciou.
- Mechanicky odolné pri namontovaní na pevný povrch.
- Ľahká integrácia do mikrovlnných obvodov.
- Napájacie vedenia a prispôbovacie prvky môžu byť vyrobené súčasne s anténovou štruktúrou [5].

Hlavné nevýhody mikropásikových antén:

- Úzka šírka pásma.
- Nízky zisk a účinnosť.
- Značné ohmické straty v napájacej štruktúre poľa.
- Sú požadované zložité napájacie štruktúry pre vysoko - výkonné polia.
- Je zložité dosiahnuť polarizačnú čistotu.
- Parazitné vyžarovanie z napájacej sústavy a kontaktov.
- Schopnosť nižšej výkonovej spracovateľnosti (približne 100W).

- Budenie povrchových vln [5].

Mikropásiková anténa môže byť napájaná rôznymi spôsobmi. Klasifikujeme ich do dvoch kategórií: kontaktné a nekontaktné. Zdroj RF je pri kontaktnej metóde napojený priamo na vyžarujúci plátok pomocou spojovacieho prvku. Spojovací prvok môže byť mikropásik alebo koaxiálny kábel. Metóda nekontaktného prenosu využíva väzbu elektromagnetického poľa energie medzi napájacím mikropásikom a vyžarujúcim plátkom. Medzi najčastejšie spôsoby napájania mikropásikových antén patria:

- Priame napájanie koaxiálnym vedením.
- Priame napájanie mikropásikovým vedením.
- Napájanie mikropásikovým vedením väzbou cez medzeru.
- Priame vsunuté napájanie mikropásikovým vedením.
- Priame asymetrické napájanie mikropásikovým vedením.
- Napájanie viazané apretúrou.
- Napájanie viazané tesnou blízkosťou mikropásikového vedenia.
- Napájanie koplanárnym vlnovodom (CPW) [5].

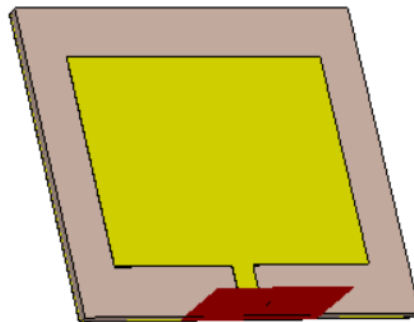
III. NÁVRH A SIMULÁCIA MIKROPÁSIKOVEJ ANTÉNY

Na základe bližšieho skúmania problematiky vytvárania mikropásikovej antény pre potreby hybridného FSO/RF systému bola navrhnutá mikropásiková anténa potencionálne použiteľná pre FSO/RF systémy. Navrhnutá mikropásiková anténa operuje v nelicencovanom frekvenčnom pásme s frekvenciou 2,4 GHz. Je založená na dielektrickom substráte typu FR04 s relatívnou permitivitou $\epsilon_r = 4,3$. Substrát sa nachádza medzi dvomi vodivými vrstvami. Pre mikropásikové antény existujú rôzne typy napájacích portov. Pre tento model bolo zvolené napájanie pomocou mikropásikového vedenia a zdroj je pripojený k tomuto portu. Parametre sú uvedené v Tabuľka 1. Mikropásikové vedenie a zemniaca plocha antény je vytvorená z medi.

Tabuľka 1 Parametre navrhutej mikropásikovej antény.

Označenie	Parameter	Hodnota
f_{range}	Rozsah pracovných frekvencií	2,0 – 2,8 GHz
W_g	Šírka substrátu a zemniacej plochy	25 mm
L_g	Dĺžka substrátu a zemniacej plochy	21 mm
H	Hrúbka substrátu	1,6 mm
W	Šírka vodivej plochy	19 mm
L	Dĺžka vodivej plochy	14,25 mm
ϵ_r	Relatívna permitivita substrátu	4,3
f_0	Pracovná frekvencia	2,4 GHz
	Typ dielektrického substrátu	FR-04

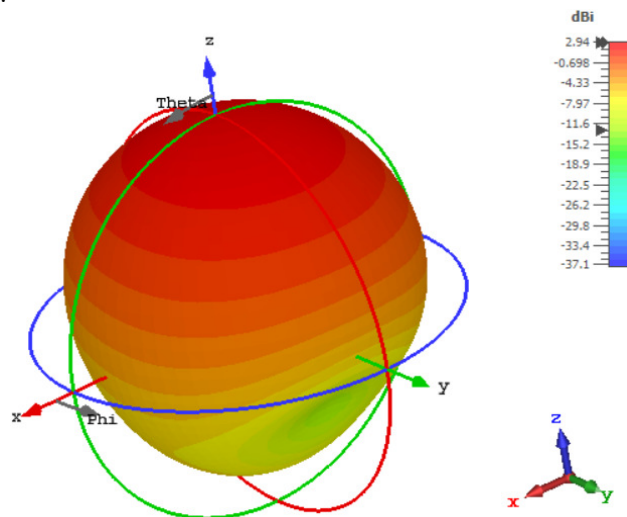
Model bol vytvorený v návrhovom prostredí CST Studio, je zobrazený na Obr. 6. Na 3D modeli je zobrazený dielektrický substrát typu FR04 slabou ružovou farbou, žltá farba predstavuje vodivú vrstvu, ktorá tvorí vodivú plochu.



Obr. 6 Model navrhovanej mikropásikovej antény.

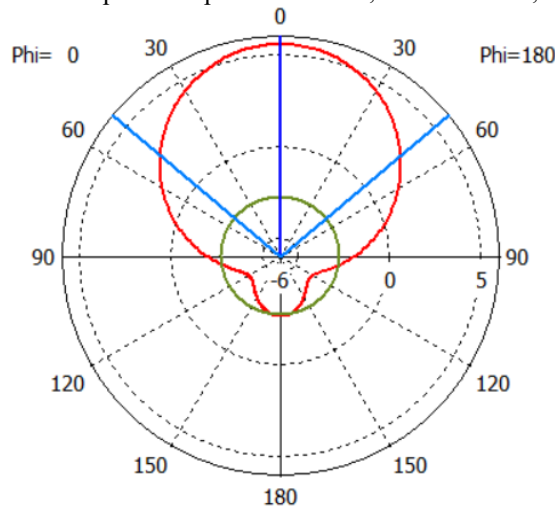
Pre tento model bola dosiahnutá hodnota maximálneho zisku 2,94 dBi. 3D vyžarovacia charakteristika je zobrazená na Obr. 7. Smerovosť hlavného laloka dosiahla hodnoty 5,56 dBi pri

frekvencií 2,4 GHz.



Obr. 7 Vyžarovacia charakteristika pre hodnotu zisku.

Hodnota celkovej efektivity pre daný model bola 0,5115 a VSWR dosiahlo hodnotu 1,62. Na Obr. 8 je zobrazená Fraunhoferova oblasť, ktorá zobrazuje vyžarovací uhol a uhlová šírka (3 dB) je 100,6 stupňov. Referenčná impedancia pri frekvencií 2,4 GHz bola 45,58 Ω .



Obr. 8 Oblasť ďalekého poľa.

IV. ZÁVER

Základnou úlohou tejto práce bolo vytvoriť vhodný koncept antény pre hybridný FSO/RF systém. Analyzovaným modelom bola mikropásová anténa s frekvenciou 2,4GHz. Tento model predstavuje základ, od ktorého sa budú následne odvíjať ďalšie návrhy. Použitý bol dielektrický substrát typu FR04. Výsledkom je návrh a analýza mikropásovej antény s pracovnou frekvenciou 2,4 GHz. Jej výhodou je nízka hmotnosť, malé rozmery, nízke výrobné náklady a ľahká integrácia do mikrovlnných obvodov. Všetky tieto výhody sú potrebné pre nasadenie tohto typu antény do hybridného FSO/RF systému. Podľa výsledkov opísaných v kapitole III. môžeme povedať, že táto anténa je pripravená na nasadenie do hybridného FSO/RF systému.

POĎAKOVANIE

Táto práca bola podporená výskumným projektom FEI-2022-84 „Techniky spracovania údajov vo vysokorýchlostných prenosových systémoch“.

LITERATÚRA

- [1] M. Khatib, „Contemporary Issues in Wireless Communications,“ 2014, 252 s., ISBN: 9535117327.
- [2] Y. Bandy, G. M. Rather, a spol., „Effect of atmospheric absorption on Millimeter Wave (mmWave) frequencies for 5G Cellular Networks,“ IET Communications, DOI: 10.1049/iet-com.2018.5044.
- [3] L. E. Frenzel, „Principles of Electronic Communication Systems,“ 2016, 945 s., ISBN: 0073373850.
- [4] J. L. Volakis, „Antenna Engineering Handbook, Fourth Edition,“ 2007, 1755 s., ISBN: 0071475745.
- [5] E. Ovseník, „Mikropásovú antény,“ dostupné na: https://data.kemt.feit.tuke.sk/EVaA_Elektromagneticke_vlny_a_anteny/_materialy/Prednasky/Pr11/Pr11_Mikropasikove%20anteny.pdf. [Citované: 18.05.2022].
- [6] C. A. Balanis, „Antenna Theory: Analysis and Design,“ 2016, 1096 s., ISBN: 1118642066.

Identifikácia rôznych typov správ o CoViD-19 pomocou zhlukovacích metód

¹Kristína MACHOVÁ, ²Natalia DANKANYCH

^{1,2} Katedra kybernetiky a umelej inteligencie, Fakulta elektrotechniky a informatiky, Technická univerzita Košice

¹kristina.machova@tuke.sk, ²natalia.dankanych@student.tuke.sk

Abstrakt — Práca sa venuje problému identifikácie rôznych typov správ o CoViD-19. Táto identifikácia sa realizuje metódou zhlukovania textových dát pomocou štandardných metrických algoritmov, ako napríklad K-means a DBSCAN. Na tento účel sa zavádza funkcia vzdialenosti medzi textami, ktorá zohľadňuje „podobnosť“ slovnej zásoby použitej v texte. V článku sa porovnávajú a analyzujú výsledky dosiahnuté rôznymi metódami zhlukovania krátkych textov.

Kľúčové slová — identifikácia rôznych typov správ, zhlukovanie textových dát, strojové učenie, K-means, DBSCAN

I. ÚVOD

Pandémia koronavírusu a s ňou spojená neistota, strach z budúcnosti nemali najlepší vplyv na duševné zdravie ľudí - úroveň úzkosti sa zvýšila. Celý svet žije v takom napätom režime už viac ako rok. Koronavírus sa stal testom nielen pre fyzické zdravie, ale aj pre duševné zdravie. Skúsenosti a obavy vyplývajú zo skutočnosti, že ľudia, ktorí sa snažia tomuto všetkému porozumieť, sú ešte viac ponorení do negatívnych informácií. Začínajú hľadať informácie na internete a snažia sa pochopiť, aké sú príznaky, ako sa vyliečiť, čo sa stane, ak ochoriete, prečo zomierajú na koronavírus. Pritom negatívne informácie prúdia doslova odkiaľkoľvek: od správ, sociálnych sietí a okamžitých správ. Počas pandémie COVID-19 sa ľudia stali o 47% závislejšími od internetu a sociálnych médií, keďže viac komunikovali spolu prostredníctvom online služieb kvôli zavedeným obmedzeniam.

Technológie sociálnych sietí bohužiaľ vedú ľudí do informačnej bubliny rovnako zmýšľajúcich ľudí v takzvanom komorovom efekte. V tejto bubline ľudia nevidia takmer žiadne alternatívne názory, čo posilňuje ich predsudky. Ak človek verí napríklad tomu, že koronavírus je údajne biologická zbraň, potom je dosť veľká pravdepodobnosť, že v neho veria aj priatelia tohto človeka. A pre takého človeka je dosť ťažké prekročiť hranice svojej informačnej bubliny, pretože informácie, ktoré vidí na svojej obrazovke, potvrdzujú jeho vieru v koronavírus, biologickú zbraň.

V súčasnosti, kedy sa zhromažďovanie rozmanitých informácií o Covid-19 prebieha obrovským tempom, sa najdôležitejšou úlohou stáva ich systematizácia a štrukturalizácia s cieľom vytvorenia čo najlepšej predstavy o nich. Jeden zo základných prístupov pre riešenie tejto úlohy pochádza už z dávnych dôb a jeho autorom je Democrita. V “liste učenenému susedovi” píše: “Ak vy, môj priateľ, potrebujete vyriešiť zložitú hromadu faktov alebo vecí, najskôr ich rozložíte na malý počet hromád podobnosti. Obrázok bude jasný a vy pochopíte podstatu týchto vecí.”. Tento spôsob – tvorba klasifikačných štruktúr zo súboru neštruktúrovaných objektov – sa nazýva zhlukovacia analýza a ma niekoľko synonym, napríklad: taxonómia, automatická klasifikácia, učenie bez učiteľa atď.

Zhlukovanie textov je možné vykonať aj pomocou pravdepodobnostných metód, napríklad pomocou pravdepodobnostnej latentnej sémantickej analýzy (PLSA) alebo latentnej Dirichletovej alokácie (LDA). V tejto práci sme sa zamerali na zhlukovanie kolekcie textov pomocou štandardných metrických algoritmov klastrovania, ako K-means alebo LSA/LSI.

Na otestovanie navrhovanej metódy klastrovania sme vykonali experimenty so textami prevzatými z Twitteru použitím rôznych metód zhlukovania. Výsledky experimentov boli

hodnotené podľa matice výsledku (Confusion Matrix), skóre homogenity zhuklu (Homogeneity Score) a hodnotenia siluety zhuklu (Silhouette Score) a napokon analyzované.

II. ZHLUKOVACIE METÓDY

Účelom zhukovania, resp. zoskupovania textov/dokumentov je automatická identifikácia skupín, sémanticky podobných textov medzi daným pevným množstvom textov. Je potrebné poznamenať, že skupiny sa tvoria iba na základe párovej podobnosti popisov textov a vopred nie sú špecifikované žiadne charakteristiky týchto skupín. Tento prístup umožňuje vytvoriť abstraktnejší model informačného poľa znížením počtu uvažovaných objektov. Namiesto skúmania každého jednotlivého zástupcu vzorky je možné zamerať sa na skúmanie tried podobných objektov, zatiaľ čo podobnosť v rámci triedy umožňuje nahradiť množiny objektov z tejto triedy určitým referenčným (ideálnym) vzorom, ktorého implementáciou sú tieto objekty. Ak je určitá trieda pre výskumníka obzvlášť zaujímavá, potom znalosti všeobecných zákonitostí a referenčných zástupcov pre ňu umožňujú rozšíriť túto triedu a hľadať všetky nové objekty, ktoré s ňou súvisia. Zhukovanie textov možno použiť na výber skupín textov rovnakého predmetu z kolekcie textov. Tieto skupiny môžu byť vhodné na identifikáciu typov ľudí (autorov textov) a formovanie vzorcov správania pre každý typ. Táto úloha súvisí s úlohami hľadania skrytých neštruktúrovaných informácií. Vzhľadom na množstvo kolekcii textov a subjektivitu vnímania témy textu čitateľom je ťažké posúdiť kvalitu zhukovania. Preto neexistuje všeobecne akceptovaná funkcia kvality zhukovania textov. V publikácii [1] sú zhukovacie algoritmy charakterizované ako typický príklad učenia bez učiteľa, teda nekontrolovaného učenia.

A. Metóda K-means

Vytvára nehierarchickú množinu zhukov, ktoré sú prezentované vo forme centroidov, ktoré sú „ťažiskom“ všetkých dokumentov zahrnutých do zhuklu. Algoritmus je založený na iteračnom procese stabilizácie klastrových centroidov. Hlavnou charakteristikou klastra je jeho centroid (mean) [2]. Algoritmus pozostáva z nasledujúcich krokov:

1. Vyberajú sa počiatočné centroidy pre viac tried. Existuje niekoľko techník:
 - a. Z množiny dokumentov sa náhodne vyberie k dokumentov.
 - b. Na základe Bayes odhadu súboru dokumentov nájdenie základnej štatistiky pre odhad a zistenie počtu zhukov a ich centroidov vhodných pre danú vzorku.
 - c. Empirické odhady počtu klastrov a ich centroidov.
2. Všetky texty kolekcie sú distribuované medzi zhuky. Dokument patrí iba do jedného zhuku, na ktorého centroid sa najviac podobá podľa danej metriky podobnosti.
3. Centroidy klastrov sa prepočítavajú na základe novej sady dokumentov v každom klastri. Ak sa ťažisko klastra pohlo, potom sa výpočtový cyklus opakuje od kroku 2. Inak, ak sa ťažisko stabilizovalo, proces zhukovania sa končí.

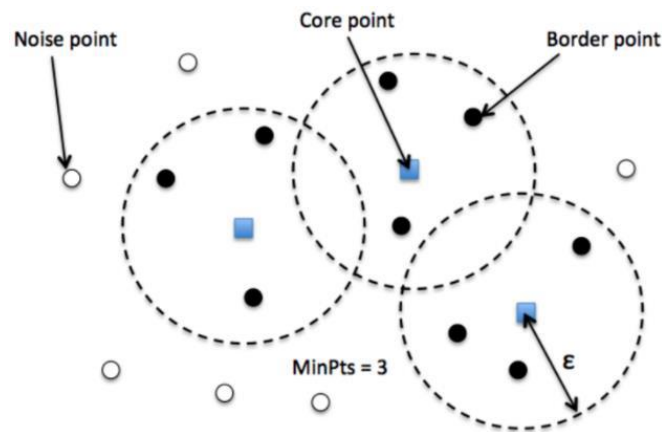
Podľa zdroja [2] má algoritmus k-means pozitíva v tom, že používa informácie z matice TF-IDF, je iba lineárne zložitý a navyše táto jednoduchá metóda môže v prípade potreby zhromažďovať informácie na ďalšie zvýšenie presnosti - použitie bayes odhadov parametrov zoskupenia. Na druhej strane má táto metóda aj nevýhody, hlavne v potrebe nastavovania počtu zhukov a taktiež v prípade náhodného výberu centroidov sa výsledky získané na tej istej vzorke dokumentov môžu líšiť.

B. Metóda DBSCAN

DBSCAN (Density-based Spatial Clustering of Applications with Noise) [3] je algoritmus na zhukovanie objektov na základe hustoty. Ak je daná množina pozorovaní - bodov v príznakovom priestore, algoritmus zoskupuje body, ktoré sú blízko seba (body s mnohými blízkymi susedmi), a označuje ako odľahlé body, ktoré sú v oblastiach s nízkou hustotou (ktorej najbližší susedia ležia ďaleko). Algoritmus DBSCAN sa často používa na vyhľadanie asociácií a štruktúr v dátach, ktoré je ťažké nájsť ručne, ale ktoré môžu byť relevantné a užitočné pri hľadaní vzorcov a predpovedaní trendov.

Tento algoritmus používa zadanú vzdialenosť na oddelenie hustých zhukov od okolitých príkladov. DBSCAN je najrýchlejšia metóda vytvárania zhukov, ale je závislá na vhodnej metrike vzdialenosti na získanie dobrých výsledkov pre všetky potenciálne zhuky v prípade, keď dáta sú charakteristické podobnou hustotou. Základnou koncepciou algoritmu DBSCAN je nájsť oblasti s vysokou hustotou, ktoré sú od seba oddelené oblastami s nízkou hustotou. Činnosť algoritmu je ilustrovaná na Obr.1. DBSCAN pracuje v nasledovných krokoch:

1. Určuje sa typ každého bodu. Každý údajový bod v našej množine údajov môže byť jeden z nasledujúcich:
 - *Základný bod*: Dátový bod je základný bod, ak sa v jeho blízkosti nachádza najmenej M bodov, to znamená v určenom polomere (R).
 - *Hraničný bod*: Údajový bod sa klasifikuje ako hraničný bod, ak: o Jeho susedstvo obsahuje menej ako M dátových bodov, alebo je prístupný z nejakého centrálného bodu, t.j. je umiestnený vo vzdialenosti R od centrálného bodu.
 - *Šumový bod*: Odľahlý bod je bod, ktorý nie je základným bodom A má menej ako M bodov vo vzdialenosti n od seba.
2. Odľahlé hodnoty sú odstránené.
3. Susedné základné body sú spojené a umiestnené v jednom klastru.
4. Každému klastru sú pridelené hraničné body.



Obr. 1 Ilustrácia práce algoritmu DBSCAN s rôznymi typmi bodov.

Podľa zdroja [3] sú výhody tejto metódy nasledovné: nie je potrebné uvádzať počet zhukov, zhukovanie efektívne zvláda odľahlé hodnoty a hlučné súbory údajov a napokon výhodou je aj to, že vyhľadáva oblasti s vysokou hustotou, ktoré sú navzájom oddelené oblasťami s nízkou hustotou. Nevýhody metódy sú, že sa vyžadujú dva parametre: polomer (R) a minimálny počet bodov (M), že nedokáže efektívne spracovať multi-dimenzionálne dáta a že zhukovanie nefunguje veľmi dobre pre riedke množiny údajov alebo pre údajové body s rôznou hustotou.

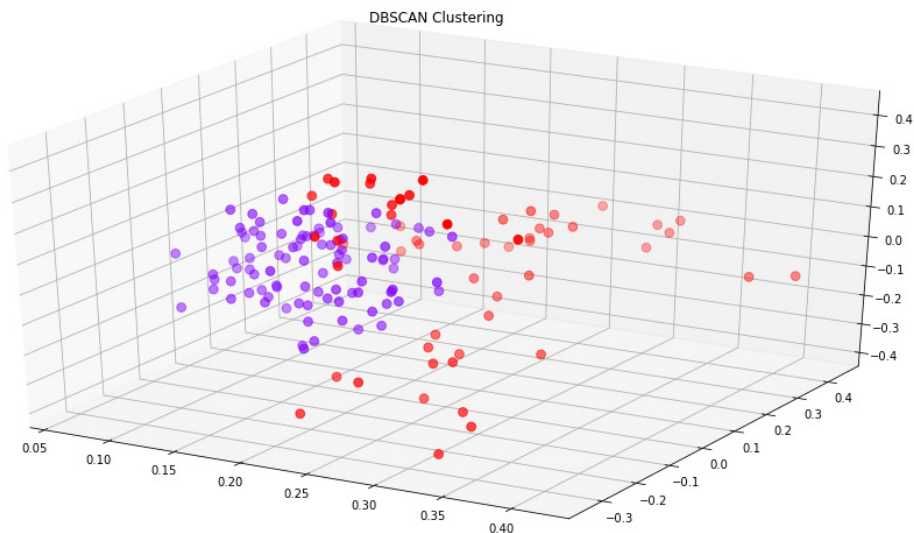
C. Výsledky testov a ich analýza

Návrh prístupu bol implementovaný v Anaconda (Python Jupiter) a v programovacom jazyku Python. Dataset obsahoval 150 tweetov. Každý tweet bol zaradený do jednej z štyroch tried: súvisiace s rúškami, umývaním rúk a sociálnou distanciou (1), súvisiace s vakcínami a očkovaním (2), súvisiace s počtom nakazených a počtom otestovaných (3) a s inou informáciou o CoVid-19 (4). V rámci predspracovania bola použitá funkcia „`tokenize_and_lemm()`“ a „`words_to_tfidf()`“. Dátová množina bola predspracovaná pomocou tokenizácie, prevodu na malé písmená, eliminácie stop-slov, lematizácie a TF-IDF indexácie.

Pre hodnotenie dosiahnutých výsledkov boli zvolené nasledovné metriky:

- *Homogeneity score (hodnotenie homogenity)* - kontrola súladu klastrového algoritmu s dôležitou požiadavkou: klaster by mal obsahovať iba vzorky patriace do rovnakej triedy
- *Silhouette score (hodnotenie siluety)* - ukazuje, ako podobný je objekt svojmu klastru v porovnaní s inými klastrami
- *Confusion Matrix (matica výsledku)*- ukazuje počet správne a nesprávne klasifikovaných príkladov v porovnaní so skutočnými výsledkami.

Porovnanie výsledkov oboch modelov je uvedené v Tabuľke 1. Z týchto výsledkov vyplýva, že metóda k-means je vhodnejšia tento typ úloh aj keď vizualizácia výsledkov modelu DBSCAN v 3D priestore ukázala sľubnosť aj tejto metódy.



Obr. 2 3D vizualizácia výsledkov modelu DBSCAN.

Tabuľka 1
 Výsledky testov zhukovacích metód K-means a DBSCAN.

Metriky hodnotenia	Presnosť trénovacej množiny [%]	Presnosť testovacej množiny [%]
Presnosť (macro avg)	0.6970	0.0978
Presnosť (weighted avg)	0.6854	0.0935
Návratnosť (macro avg)	0.6866	0.1667
Návratnosť (weighted avg)	0.68	0.1781
Miera F1 (macro avg)	0.6910	0.1178
Miera F1 (weighted avg)	0.6820	0.1240
Správnosť	0.68	0.17
Hodnotenie homogenity	0.4459	0.1523
Hodnotenie siluetu	0.0135	0.0039
Čas spracovania	114 ms	275 ms

III. ZÁVER

V článku je uvedený popis riešeného problému a zhukovacích metód, ktoré boli zvolené pre implementáciu modelov zhukovania. Cieľom tejto práce bolo vytvoriť zhukovací model, ktorý dokáže identifikovať rôzne typy správ o CoViD-19 pomocou zhukovacích metód. Článok uvádza výsledky metrik hodnotenia výsledných modelov. Výsledkom hodnotenia rôznych algoritmov pre klastrovanie dokumentov bolo, že pri probléme rozdelenia množstva záznamov s informáciami o CoViD-19 na klastre najlepšie výsledky ukázal algoritmus K-means.

POĎAKOVANIE

Práca prezentovaná v tomto článku bola podporená Slovenskou agentúrou na podporu vedy a výskumu v rámci projektu VEGA 1/0685/21.

LITERATÚRA

- [1] MACHOVA, K. Strojové učenie v systémoch spracovania informácií. Elfa s.r.o, Košice, 2009, 85s, ISBN 978-80-8086-130-8.
- [2] PELLEG D., MOORE A., 2002. *X-means: Extending K-means with Efficient Estimation of the Number of Clusters*. Pittsburgh, PA 15213 USA. 2002, Dostupné na: https://www.researchgate.net/publication/2532744_X-means_Extending_K-means_with_Efficient_Estimation_of_the_Number_of_Clusters [cit. 28.4.2022].
- [3] ESTER M., KRIEGEL H.-P., Sander J., XU X. *A density-based algorithm for discovering clusters in large spatial databases with noise*. Proc. 2nd Int. Conf. on Knowledge Discovery and Data Mining (KDD-96), Germany, AAAI Press, 1996.

Využitie algoritmu počítačového videnia v priemyselnej aplikácii

¹Vladimír JACKO, ²Branislav MADOŠ, ³Anton BALÁŽ

^{1,2,3}Department of Computers and Informatics, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Slovak Republic

¹vladimir.jacko@student.tuke.sk, ²branislav.mados@tuke.sk, ³anton.balaz@tuke.sk

Abstrakt – Článok sa zaoberá návrhom algoritmu umožňujúceho, vďaka aplikácii algoritmov počítačového videnia, riešiť problém vyplývajúci z priemyselnej praxe, ktorým je potreba spočítať kovové rúrky uložené v prepravníku. V úvodnej časti sú predstavené spôsoby predspracovania a segmentácie obrazu ako aj algoritmus Houghovej transformácie. V ďalšej časti článok predstavuje navrhnutý algoritmus, ktorý nielen jednoducho identifikuje objekty v obraze, ale dopĺňa aj dopočítavanie rúrok, ktoré sú v obraze skryté. Na tento účel boli využité poznatky z teórie grafov. Implementované riešenie bolo otestované na poskytnutých testovacích dátach a výsledky porovnané s komerčne dostupným softvérom CountThings.

Keywords – Houghova transformácia, počítačové videnie, predspracovanie obrazu, segmentácia obrazu

I. ÚVOD

Počítačové videnie je inžinierska disciplína, ktorej cieľom je budovať algoritmy, ktoré dokážu to, čo človek: rozpoznať a extrahovať užitočné informácie z obrazov. Tento problém však nie je neriešiteľný. Nápodvedou je nám naše vlastné, ľudské videnie. Človek dokáže bez problémov na obrázku spočítať stromy, autá či iné objekty aj keď nie sú celé v zábere či kompozícia scény nie je ideálna. Počítače sú v súčasnosti schopné zachytiť obraz v kvalite a rozlíšení vyššom ako ľudské oko. Avšak tak, ako je rozdiel medzi počutím a porozumením, rovnako je rozdiel vo videní a skutočnom extrahovaní informácií z obrazu.

II. SÚVISIACA PRÁCA

V tejto sekcii článku sú stručne priblížené techniky predspracovania, segmentácie obrazu a Houghovej transformácie.

A. Predspracovanie obrazu

Predspracovanie obrazu je proces, ktorého cieľom je zlepšenie kvality digitálneho obrazu. Je dôležitým krokom pri spracovaní veľkých dátových sád, výsledkom sú vstupné dáta vhodnejšie na ďalšie spracovanie.

Hodnota pixelov v obraze získaná vstupnými zariadeniami, ktorá neodráža skutočnú intenzitu zo skutočnej scény, vytvára v digitálnom obraze šum. Na obnovenie pôvodného obrazu je ho potrebné odstrániť. Preto bol zavedený koncept filtrovania obrazu. Šum sa na fotografiách prejavuje ako zmenie alebo jasné farebné škvrny [1]. Za jeho vznikom často stoja nepriaznivé svetelné podmienky počas zachytávania obrazu alebo počas prenášania obrazu po sieti. Okrem zjavných negatívnych efektov na kvalitu obrazu zvyšuje dátový objem súborov [2].

Hornopriepustný filter je využívaný na zvýšenie kontrastu obrazu. Aj jednoduché potlačenie rozostrenia obrazu možno považovať za určitý druh hornopriepustnej filtrácie. Tento druh filtrácie je založený na konvolúcii masky s obrazom, preto je kritický správny výber masky.

Dolnopriepustný filter, napríklad Gaussov filter, je inverznou operáciou hornopriepustného filtra popísaného vyššie. Kým cieľom hornopriepustnej filtrácie bolo obraz zostríť, dolnopriepustný filter obraz naopak rozmaže. Spôsob výpočtu je totožný, avšak výsledok nie je použitím hornopriepustného filtra zvrátiteľný. Dolnopriepustný filter nezvratne modifikuje obraz.

Mediánový filter bol predstavený v 70-tych rokoch minulého storočia Johnom Tukeyom na vyrovňovanie diskretných dát v jeho knihe Exploratory Data Analysis [3]. Keďže mediánový filter

efektívne tlmí impulzný šum pri zachovaní hrán, je často využívanou technikou predspracovania obrazu [4].

Mediánový filter najprv zoradí intenzity susedných pixelov vstupného pixelu vo vzostupnom poradí podľa veľkosti jas. Potom nastaví intenzitu výstupného pixelu rovnajúcu sa štatistickému mediánu tohto triedeného kernelu. Pre dosiahnutie lepších výsledkov je možné proces opakovať. Výsledný obraz zvykne byť kvalitnejší v porovnaní s hornopriepustnými a dolnopriepustnými filtermi. Funguje dobre pre rôzne typy šumu, s menším rozmazaním ako lineárne filtre podobnej veľkosti [2][5].

B. Detekcia hrán

V disciplíne počítačového videnia, detekcia hrán je procesom, ktorého cieľom je extrahovať obrysy objektov zachytených v obraze. Metódy, ktoré sem radíme, pracujú na princípe analyzovania rozdielu úrovne jas, ktoré sú pozorovateľné v obraze medzi dvoma rôznymi objektami, alebo pri prechode z objektu na pozadie [6]. Cieľom detekcie hrán je lokalizovanie týchto rozdielov a identifikácia fyzikálnych fenoménov, ktoré ich spôsobujú [7].

Detektor hrán Canny je operátor detekcie hrán, využívajúci viacstupňový algoritmus na detekciu širokého rozsahu hrán v obrazoch. Autorom je John F. Canny a svoj postup prezentoval v článku [8] v roku 1986. Cieľom bolo priniesť optimálny algoritmus detekcie hrán [9]. Canny sformuloval tri požiadavky na optimálny detektor hrán:

- 1) Minimálna chybovosť: Pravdepodobnosť detekcie pravej hrany musí byť maximalizovaná kým pravdepodobnosť detekcie falošnej hrany musí byť minimalizovaná.
- 2) Presnosť: Poloha hrany musí byť určená s čo najväčšou presnosťou, teda čo najbližšie k reálnej hrane.
- 3) Jednoznačnosť: Jedna hrana musí byť zachytená detektorom práve jedenkrát, nesmie dochádzať k duplicitám [10].

Postup Cannyho detekcie hrán sa vykonáva v piatich krokoch:

- 1) Potlačenie šumu použitím dolnopriepustnej filtrácie, napríklad Gaussovým filtrom.
- 2) Detekcia horizontálnych a vertikálnych hrán použitím Sobelovej filtrácie.
- 3) Stenčovanie hrán - redukcia hrúbky hrán na jeden pixel hľadáním lokálnych maxim gradientu.
- 4) Dvojnásobné prahovanie - určenie kandidátov na hranice oblastí. Algoritmus pracuje s dvomi zvolenými hodnotami pre gradientný obraz:
 - a) Ak hodnota gradientu pixelu je vyššia ako väčšia zvolená prahová hodnota, pixel je považovaný za silnú hranu.
 - b) Ak hodnota gradientu pixelu je nižšia ako menšia zvolená prahová hodnota, pixel je považovaný za slabú hranu.
- 5) Spájanie hrán - finálne hranice sú určené pospájaním silných hrán a všetkých slabých hrán s nimi spojených [5][9].

Houghova transformácia

Houghova transformácia sa používa na detekciu parametricky opísateľných tvarov a kriviek v obrazovej analýze, počítačovom videní a digitálnom spracovaní obrazu. Účelom tejto transformácie je nájsť objekty, teda analytický opis hľadaných tvarov a kriviek v obraze. Tento proces prebieha formou hlasovania v parametrickom priestore hľadaných kriviek na základe najväčšej tvarovej zhody [5]. Vo všeobecnosti platí, že výpočet a veľkosť poľa akumulátora sa polynomiálne zväčšujú s počtom parametrov. Základná Houghova technika tu opísaná je teda praktická len pre jednoduché krivky [11].

Na detekciu priamky v obraze je nutné nájsť čo najviac bodov v obraze, ktoré spĺňajú jej predpis a v priestore parametrov majú niekoľkonásobný priesečník. Pri prehľadávaní bodov obrázka sa potrebujeme obmedziť na body, ktoré reprezentujú hrany, resp. prechody medzi jednotlivými objektami. Práve tieto body obrázka ležia na hľadanej priamke, ktorú chceme nájsť a majú viacnásobný priesečník v parametrickom priestore [5]. Pred spracovaním je praktické vstupný obraz predspracovať filtrovaním, najčastejšie sa využíva Cannyho alebo Sobelov filter.

Hľadané priamky je možné nájsť pomocou priesečníkov v priestore parametrov, kde každá dvojica čiar predstavuje hranový bod obrazu. Toto vyhľadávanie je realizované formou hlasovania. Na uchovanie hodnôt hlasovania sa využíva akumulátor. Akumulátor je pomocný priestor diskretných parametrov (q,k), ktorý možno reprezentovať ako maticu:

$$\begin{bmatrix} k_{\min}' q_{\min}' & \dots & k_{\min}' q_{\max} \\ \dots & & \dots \\ k_{\min}' q_{\min}' & \dots & k_{\min}' q_{\max} \end{bmatrix}$$

Pre každý hranový bod (x, y) obrazu splňajúci rovnicu

$$q = -xk + y \quad (1)$$

inkrementujeme príslušnú hodnotu parametra (k, q) v akumulátore. Extrémne hodnoty uložené v akumulátore tak zodpovedajú parametrom nájdených priamok v obraze. Pri zvislých čiarach sa však hodnoty parametrov k a q rozchádzajú do nekonečna. Toto vyjadrenie sa preto využíva len pre hľadanie priamok, ktoré nie sú zvislé. Tento problém vyriešil Duda a Hart vo svojej práci *Pattern Classification*[12] vyjadrením priamky v polárnych súradniciach:

$$y = -\frac{\cos\theta}{\sin\theta}x + \frac{r}{\sin\theta} \quad (2)$$

Výraz upravíme do tvaru

$$r = x \cos\theta + y \sin\theta \quad (3)$$

pričom r je vzdialenosť normály priamky od bodu $[0,0]$ súradnicového systému a θ vyjadruje uhol medzi touto normálou a kladnou časťou x-ovej osi.

Po zovšeobecnení Houghovej transformácie sa nám otvárajú možnosti rozšíriť detekciu priamok na zložitejšie útvary, akými sú napríklad parametrické krivky. Najznámejšie z nich sú kružnica či elipsa.

Kružnica má v karteziánskej súradnicovej sústave všeobecný predpis

$$(x - x_0)^2 + (y - y_0)^2 = \rho^2 \quad (4)$$

Parametrické rovnice kružnice sú

$$\begin{aligned} x &= x_0 + \rho \times \cos(\phi) \\ y &= y_0 + \rho \times \sin(\phi) \end{aligned} \quad (5)$$

Pri rovnici priamky sme nepoznali dva parametre k a q . V predpise kružnice vidíme, že sú už tri: x_0 , y_0 a ρ . Parametrický priestor - akumulátor sa preto stáva trojrozmerný, čo spôsobuje zvýšenie náročnosti výpočtov. Bežne sa v praxi využíva obmedzenie veľkosti polomeru kružníc r , ten sa pohybuje v určenom rozmedzí. Presnosť odhadu závisí od použitého obrazu a veľkosti kružníc, ktoré obsahuje. V prípade, že by sme vynechali spomínané obmedzenie, výrazne to zvýši výpočtovú náročnosť algoritmu. Postup hľadania kružníc v obraze môžeme vyjadriť nasledovným pseudokódom:

```

Načítanie obrazu
Aplikácia hranového filtra
Akumulátor = 0
for all x do
  for all y do
    for all  $x_0$  do
      for all  $y_0$  do
         $r = \text{sqrt}((x - x_0)^2 - (y - y_0)^2)$ 
        akumulátor inkrementuj ( $x_0, y_0, r$ )
      end for
    end for
  end for
end for
end for

```

III. NÁVRH RIEŠENIA

Táto sekcia článku sa zaoberá zhrnutím návrhu riešenia, ktoré má za úlohu spočítať rúrky umiestnené na dopravníku. Postupne vysvetlí a demonštruje použité postupy na predspracovanie obrazu, identifikáciu kružníc v obraze a algoritmus na dopočítanie rúrok, ktoré nie je možné identifikovať metódami počítačového videnia.

A. Predspracovanie vstupných dát

Po experimentoch s rôznymi technikami predspracovania, prahovania a filtovania boli najlepšie výsledky dosiahnuté po prevedení obrazu do škály sivej a následnom aplikovaní mediánového filtra.

```

cvtColor(src, gray, COLOR_BGR2GRAY);
medianBlur(gray, gray, 7);

```

B. Identifikácia rúrok

Pre identifikáciu rúrok v obraze bola použitá Houghova transformácia. V knižnici jazyka C++ OpenCV je dostupná funkcia `HoughCircles`, ktorá vyžaduje na vstupe deväť parametrov, bližšie popísaných v dokumentácii knižnice. Na Obr. 1 je znázornený výsledok funkcie `HoughCircles` za použitia nasledujúcich parametrov:

- `img`: vstupný obraz
- `method`: `HOUGH_GRADIENT`
- `dp`: 1
- `min_dist`: 21
- `param_1`: 100
- `param_2`: 21
- `min_radius`: 11
- `max_radius`: 20

Počet identifikovaných kružníc je veľkosťou výstupného vektora kružníc `circles`, teda `circles.size()`.

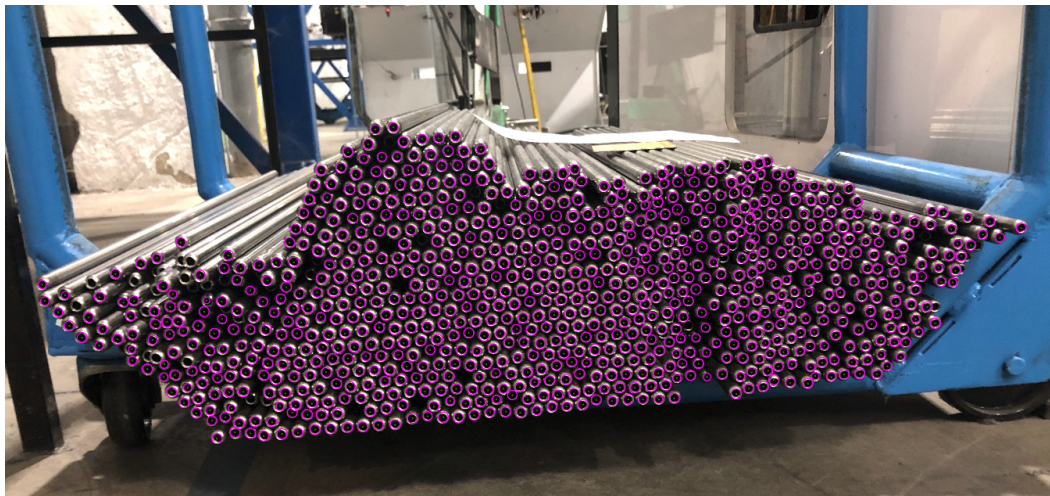


Fig. 1 Identifikácia rúrok v obraze.

C. Dodatočná detekcia predpokladaných rúrok

V špecifickejšej doméne riešeného zadania, rátania rúrok uložených na hromade, vzniká problém, že nie všetky rúrky sú v obraze viditeľné. V závislosti od úhľadnosti uloženia hromady sú niektoré rúrky vtlačené hlbšie alebo naopak vytiahnuté dopredu. Rúrky vtlačené dnu sú čiastočne alebo celkom neviditeľné v obraze. Do určitej miery sa dá tento problém kompenzovať technikami predspracovania obrazu a prahovaním, avšak nie úplne. Vtlačená rúrka v obraze sa prejaví ako tmavé miesto obkolesené zo všetkých strán detegovanými rúrkami. Na základe počtu detegovaných rúrok ohraničujúcich takéto tmavé miesto je možné predpokladať počet vtlačených rúrok v danom mieste. Identifikácia vtlačených rúrok bola preto riešená ako problém nájdenia cyklov v grafe, ktorý vznikne pospájaním stredov kružníc identifikujúcich rúrky. Dĺžku cyklu je možné následne transformovať do počtu neidentifikovaných rúrok.

Pri vytváraní grafu, teda spájaní stredov kružníc, algoritmus pracuje s nameranou hodnotou priemeru kružníc. Na prvý pohľad sa zdá samozrejme, že dĺžka každej hrany má byť presne polomer krát dva. Pozorovaním však bolo zistené, že to nie je vždy pravda a napriek tomu, že všetky rúrky majú rovnaký polomer, namerané hodnoty majú pomerne značnú odchýlku. Výpočet maximálnej dĺžky hrany komplikuje mnoho faktorov, medzi hlavné patria:

- 1) Uhol, pod ktorým je konkrétna rúrka zachytená na obraze.
- 2) Hrúbka stien rúrok.
- 3) Zarovnanie rúrok v hromade.

Z toho dôvodu bol pridaný ďalší vstupný parameter, ktorým sa násobí súčet polomerov detegovaných kružníc. Výsledný vzorec na výpočet maximálnej dĺžky hrany je preto:

```
float max_edge_length = (r + neigh_r) * argv[8]; , kde r je polomer  
kružnice, pre ktorú hľadáme susedné vrcholy, neigh_r je polomer susednej kružnice a
```


`argv[8]` je vstupný parameter. Tento parameter je typu `float` a jeho odporúčaná hodnota je 1,5 až 3.

Algoritmus prechádza poľom `circles`, ktoré obsahuje x a y súradnice stredu a polomer pre každú detegovanú kružnicu. Pre každú kružnicu sú vyhľadávané všetky kružnice, ktorých stredu sú vzdialené menej ako `max_edge_length`. Ak je nájdená vyhovujúca dvojica stredu kružníc, je vykreslená úsečka medzi týmito dvoma bodmi a táto hrana je pridaná do poľa vektorov grafu `gr`. Výstupom je spracovaný obraz, ktorý možno vidieť na Obr. 2.

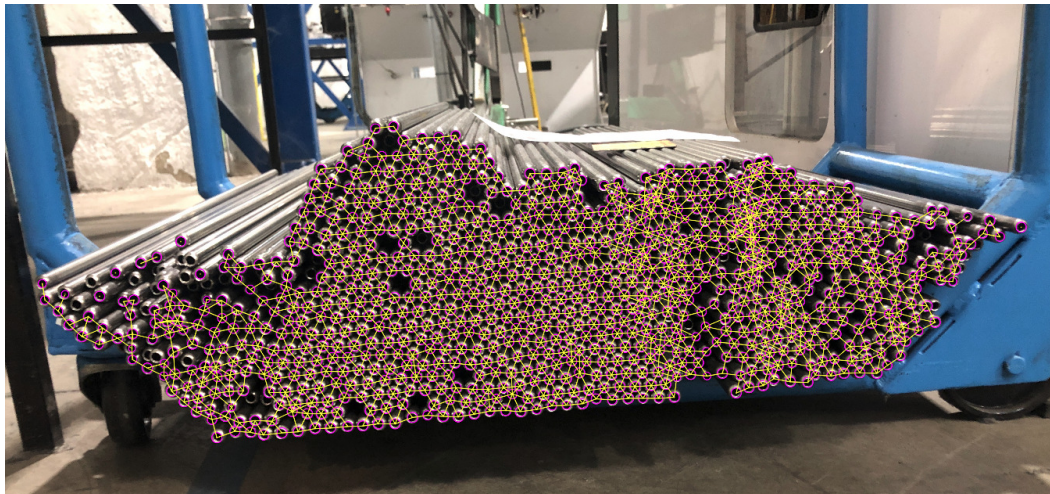


Fig. 2 Vytvorenie grafu zo stredov detegovaných kružníc v obrázku 1.

Na detekciu vyhovujúcich cyklov v grafe bola implementovaná funkcia `chordles_cycle`. Jej jediným vstupným parametrom je celočíselná hodnota počtu vrcholov v grafe. Tento postup je inšpirovaný Dijkstrovým algoritmom na nájdenie najkratšej cesty v orientovanom grafe založenom na prehľadávaní do hĺbky (v angličtine depth-first search, DFS).

- 1) Pre každý vrchol grafu sa nájdu všetky cykly, ktoré tento vrchol obsahujú.
- 2) Odfiltrujú sa cykly kratšie ako 5 vrcholov, pretože sú príliš malé na to, aby obsahovali skrytú rúrku.
- 3) Odfiltrujú sa cykly dlhšie ako 12 vrcholov, pretože pri väčšej veľkosti cyklu sa už nedá spoľahlivo odhadovať súvislosť s počtom skrytých rúrok.
- 4) Všetky vrcholy cyklu sa uložia do mapy `nodes_map`.
- 5) Overí sa, že cyklus neobsahuje tetivy funkciou `no_inner_edges`.
- 6) Overí sa, že cyklus neobsahuje vnútorný vrchol funkciou `no_inner_vertex`.
- 7) Ak cyklus s týmito vrcholmi nebol doteraz nájdený, je považovaný za dôkaz o existencii skrytých rúrok. Počítadlo skrytých rúrok `hidden_tubes` sa navýši podľa dĺžky cyklu.

Výsledný spracovaný obraz je zobrazený na Obr. 3. Obvodové vrcholy dutín, v ktorých boli identifikované skryté rúrky, sú zvýraznené bielou farbou.

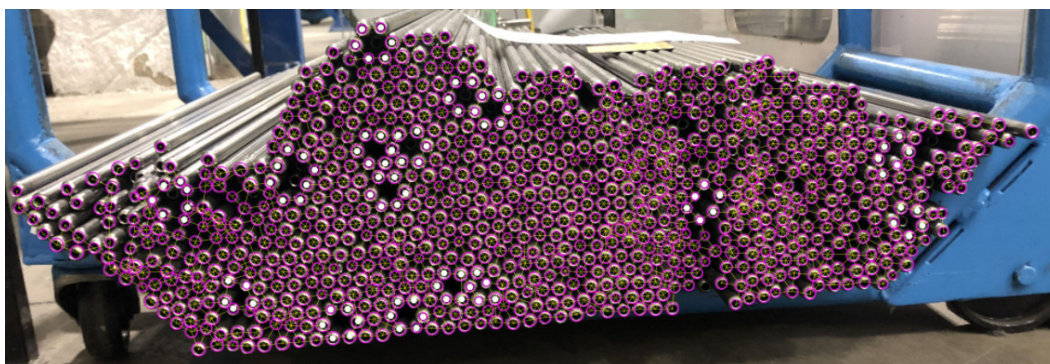


Fig. 3 Výsledok spracovania.

Table 1
Sumár výsledkov na testovacích obrazoch.

	Skutočný počet	Navrhnutý algoritmus	CountThings
Testovací obraz č. 1	1166	1151	1140
Testovací obraz č. 2	1292	1284	1262
Testovací obraz č. 3	655	653	646
Testovací obraz č. 4	1298	1264	1272
Testovací obraz č. 5	1380	1359	1374
Testovací obraz č. 6	460	450	451
Testovací obraz č. 7	862	846	834

IV. DOSIAHNUTÉ VÝSLEDKY

Na testovanie implementovaného programu boli využité testovacie fotografie z reálneho prostredia výrobného závodu. Fotografie boli na porovnanie spracované aplikáciou CountThings, ktorá sa špecializuje na počítanie objektov vo fotografiách za použitia učiacich sa algoritmov, ktoré je možné zjednotene nazvať ako algoritmy umelej inteligencie. Z testovaných dát bolo zvolených sedem obrazov na demonštrovanie výhod, nevýhod, pozitív a problémov programu navrhnutého v tejto práci. Výsledky testovania ukázali, že algoritmus navrhnutý v tejto práci dosahuje úspešnosť 98,195% kým spracovanie cez aplikáciu CountThings úspešnosť 97,975%. Navrhnutý algoritmus teda na testovacích dátach dosiahol vyššiu úspešnosť. Prehľad všetkých nameraných výsledkov je zobrazený v Tab. 1 a vizualizovaný v grafe na Obr. 4.

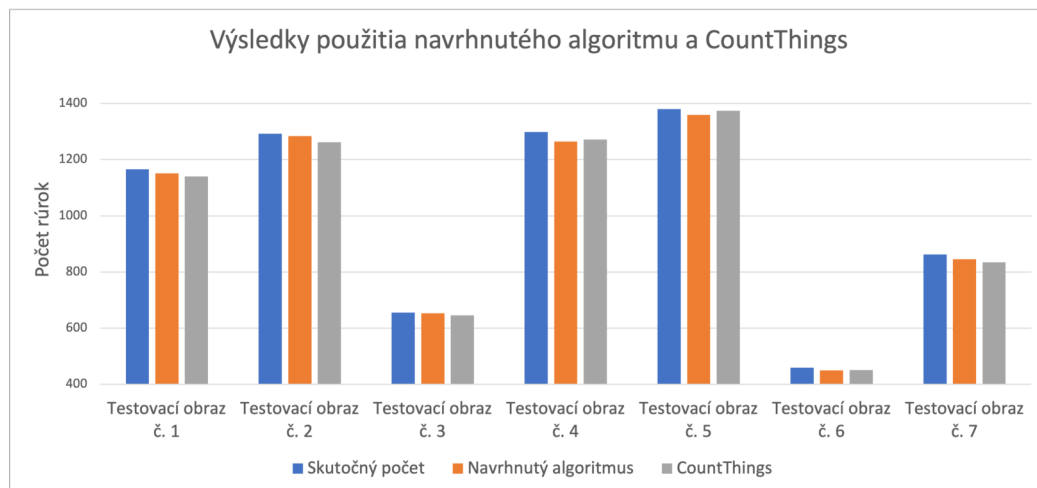


Fig. 4 Graf výsledkov použitia navrhnutého algoritmu oproti CountThings.

V. CONCLUSION

Článok prezentoval spôsob využitia techniky predspracovania, segmentácie obrazu, Houghovej transformácie na riešenie konkrétneho problému z priemyselnej praxe. Pri skúmaní poskytnutých testovacích dát vyšlo najavo, že samotná identifikácia kružníc nebude dostatočným spôsobom na spočítanie rúrok v hromade. Pri ukladaní rúrok v hromadách vznikajú dutiny, kde je jedna alebo viac rúrok vtlačených hlbšie medzi ostatné. Tieto rúrky nie sú na fotografiách viditeľné a preto nemôžu byť ani identifikované ako kružnice. Z toho dôvodu by neboli započítané do súčtu. Výsledkom riešenia problému bolo navrhnutie algoritmu, ktorý dáta z obrazu prevedie do grafu a následne je schopný identifikovať miesta, v ktorých sa s vysokou pravdepodobnosťou skrýva rúrka. Výsledky dosiahnuté na testovacích dátach boli porovnávané s výsledkami aplikácie CountThings špecializujúcej sa na identifikáciu objektov z fotografií pomocou strojového učenia. Navrhnutý algoritmus dosiahol lepšie výsledky a to najmä vďaka schopnosti dopočítavať rúrky, ktoré neboli schopný identifikovať pomocou algoritmov počítačového videnia.

ACKNOWLEDGMENT

This publication has been published with the support of the Operational Program Integrated Infrastructure within project: Research in the SANET Network and Possibilities of Its Further Use and Development (ITMS code: 313011W988), co-financed by the ERDF. This research was supported by KEGA 002TUKE-4/2021 Implementation of Modern Methods and Education Forms in the Area of Cybersecurity Towards Requirements of Labour Market.

REFERENCES

- [1] R. Chandel and G. Gupta, "Image filtering algorithms and techniques: A review," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 3, no. 10, 2013.
- [2] N. Patel, A. Shah, M. Mistry, K. Dangarwala *et al.*, "A study of digital image filtering techniques in spatial image processing," in *Proceedings of the 2014 International Conference on Convergence of Technology (I2CT)*, 2014, pp. 1–6.
- [3] J. W. Tukey *et al.*, *Exploratory data analysis*. Reading, MA, 1977, vol. 2.
- [4] G. Sicuranza, *Nonlinear image processing*. Elsevier, 2000.
- [5] J. Pavlovičová and M. Loderer, *Číslíkové spracovanie obrazu*, 1st ed. FELIA s.r.o., 2016, vol. 100ks.
- [6] J. S. Lim, *Two-dimensional signal and image processing*. Prentice-Hall, 1990.
- [7] D. Ziou, S. Tabbone *et al.*, "Edge detection techniques-an overview," *Pattern Recognition and Image Analysis C/C of Raspoznaniye Obrazov I Analiz Izobrazhenii*, vol. 8, pp. 537–559, 1998.
- [8] J. Canny, "A computational approach to edge detection," *IEEE Transactions on pattern analysis and machine intelligence*, no. 6, pp. 679–698, 1986.
- [9] X. He, J. Li, D. Wei, W. Jia, and Q. Wu, "Canny edge detection on a virtual hexagonal image structure," in *2009 Joint Conferences on Pervasive Computing (JCPC)*, 2009, pp. 167–172.
- [10] P. Bao, L. Zhang, and X. Wu, "Canny edge detection enhancement by scale multiplication," *IEEE transactions on pattern analysis and machine intelligence*, vol. 27, no. 9, pp. 1485–1490, 2005.
- [11] R. Fisher, *HIPR: Hypermedia Image Processing Reference*. John Wiley, 1996.
- [12] R. Duda, P. Hart, P. Hart, P. Hart, D. Stork, E. Library, and J. W. . Sons, *Pattern Classification*, ser. A Wiley-interscience publication. Wiley, 2001, no. diel 1.

Technológia Blockchain a smerovací protokol AODV v MANET

¹Erik SEIDENGLANZ, ²Ján PAPAJ

¹Katedra technológií v elektronike, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská Republika

²Katedra elektroniky a multimediálnych telekomunikácií, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská Republika

¹erik.seidenglanz@student.tuke.sk, ²jan.papaj@tuke.sk

Abstract – Mobilné viacpreskokové ad-hoc siete (MANET) hrajú v súčasnom období kľúčovú úlohu pri poskytovaní komunikačných možností hlavne v prípade krízových situácií. Do popredia sa dostáva otázka ako ich bude možné využiť v sieti 6. generácie. Nevyhnutným elementom úspešnej integrácie MANET do 6G sietí bude využívanie technológie Blockchain. Len nedávno sa začali skúmať vlastnosti Blockchainu a ich využitie v mobilných sieťach. Tento článok sa venuje základnej problematike implementácie technológie Blockchain do smerovania v smerovacom protokole BAODV v prostredí simulátora NS-3. Cieľom článku je skúmať vplyv technológie blockchain na samotný proces smerovania.

Keywords – AODV, Blockchain, MANET, odolnosť, smerovanie

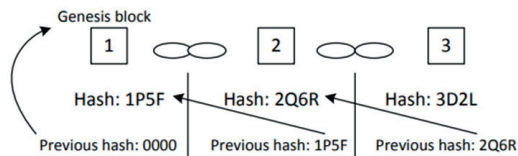
I. ÚVOD

Mobilné viacpreskokové siete (MANET) sú autonómne samo-organizované siete bez pevnej infraštruktúry. V takejto sieti funguje každý uzol súčasne ako zdrojový, smerovací a cieľový uzol. Všetky uzly v sieti sú si navzájom ekvivalentné a môžu sa voľne pohybovať, alebo sa k tejto sieti pripájať [1], [2]. Uzly, ktoré sú vo vzájomnom rádiovom dosahu, môžu priamo komunikovať a preniesť potrebné informácie. Všetky uzly siete majú bezdrôtové rozhranie na komunikáciu s iným uzlom v rádiovom rozsahu. Tento druh siete je plne distribuovaný a môže fungovať na akomkoľvek mieste bez pomoci akejkoľvek pevnej infraštruktúry (prístupové body alebo základňové stanice). Ak uzol opustí sieť a spôsobí prerušenie spojenia, postihnuté uzly môžu ľahko požiadať o nové trasy a problém sa vyrieši. To mierne zvýši oneskorenie, ale sieť bude stále funkčná.

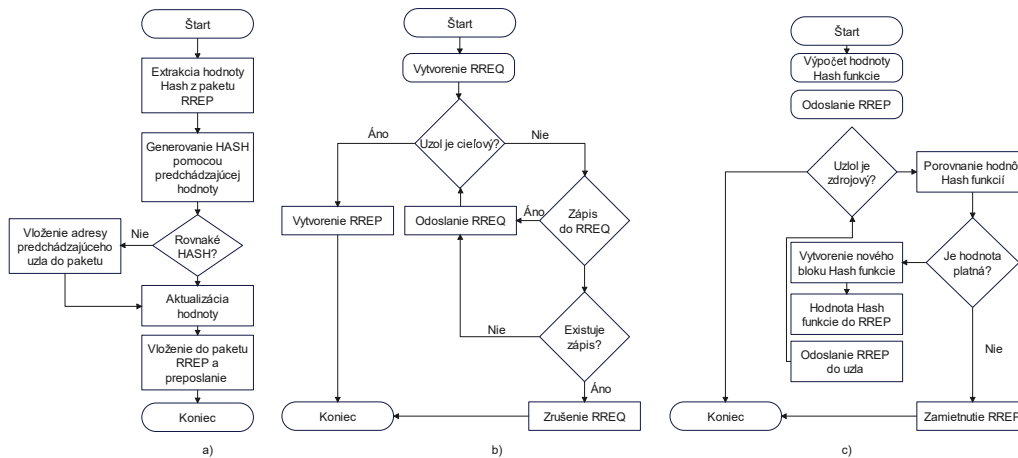
Poskytovanie odolného smerovania v bezdrôtových sieťach, ako sú MANET siete, je náročná a zložitá záležitosť z dôvodu mobilného charakteru týchto sietí. Technológia Blockchain (BC) poskytuje riešenie, ktoré spočíva v autentifikácii a konektivite k MANET sieťam [3], [4].

Technológia Blockchain (BC) sa bežne spája s kryptomenami ako napríklad Bitcoin alebo ethereum. Je to databáza záznamov transakcií, ktoré sú distribuované, overované a udržiavané sieťou počítačov okolo sveta [5]. Namiesto jediného ústredného orgánu, akým je napríklad banka, dohliada na evidenciu veľká komunita. V porovnaní s konvenčnou centralizovanou databázou informácie nemôžu byť manipulované vďaka charakteristickej štruktúre BC. Inými slovami, keď je centralizovaná databáza umiestnená na serveri, BC je distribuovaná medzi používateľmi softvéru. BC umožňuje komukoľvek v sieti prístup k záznamom všetkých ostatných, ktoré sú nemenné a znemožňujú jednej centrálnej osobe získať kontrolu nad sieťou. Keď niekto vykoná transakciu, táto transakcia ide do siete a následne počítačové algoritmy overujú jej pravosť. Akonáhle transakciu overia, pripojí sa na predchádzajúcu transakciu, tvoriac tak reťazec transakcií nazývaný Blockchain. BC je založená na decentralizovanej sieti, čo znamená, že funguje ako sieť s rovnocennými uzlami. Jednou z populárnych BC je kryptomena Bitcoin. Bitcoin poskytuje platformu na ťaženie, ukladanie a obchodovanie vďaka komplexným počítačovým algoritmom, ktoré sú naviazané na distribuovanú sieť. BC nemusí byť použitá len na transakcie, môže sa využiť aj ako register a inventár majetku [6], [7]. Kľúčové výhody BC sú decentralizácia, odolnosť, anonymita, autentickosť. Existuje viacero druhov BC, ktoré sú založené na použití, a majú nasledovné rozlišovacie znaky:

- **Verejné BC** - sú bezpečné aj napriek tomu, že sa ktokoľvek môže pridať do verejnej BC a odoslať transakciu ako blok. Tieto bloky sú overované pomocou algoritmov obsiahnutých v BC.



Obr. 1: Princíp technológie BC v MANET [4]



Obr. 2: BAODV protokol: a) Základná idea činnosti, b) Hľadanie cesty - RREQ, c) Výber cesty - RREP [3]

- **Súkromné BC** - bezpečnejší a efektívnejší ako verejná BC, pretože iba jej známi členovia sa pripájajú k sieti. Tá je zabezpečená hodnotou hašovacej funkcie a konsenzom účastníkov rovnakým spôsobom ako verejná BC.
- **BC využívajúce spoluúčasť** - ako súkromná BC, ktorá je určená pre viaceré organizácie. Do siete sa môžu pripojiť a udržiavať ju iba pozvané a dôveryhodné osoby.

BC je teda reťazec blokov, kde každý blok je braný ako transakcia na celom reťazci, z ktorého tieto dané bloky nemôžu byť odobraté, ale iba pridávané. Každý blok prítomný v reťazci pozostáva z údajov v bloku (Obr. 1). Blok obsahuje hodnoty hašovacej funkcie, ktorá sa odkazuje na tieto konkrétne údaje na konkrétnom bloku, a predchádzajúcej hodnoty hašovacej funkcie predchádzajúceho bloku. Údaje zaznamenané v BC závisia od typu danej BC [7].

A. Blockchain v MANET a BAODV

Úspešné realizovanie hodnoty hašovacej funkcie na prepojenie údajov a zdieľanie prepojených dát s uzlami v sieti BC ponúka autentifikáciu pre viacpreskokové siete. Nový vytvorený blok má aktuálne informácie o všetkých ostatných blokoch v reťazci. Tieto bloky sú chránené digitálnym podpisom dôveryhodných uzlov, ktoré riadia a zabezpečujú systém proti útokom a zlomyseľným uzlom. Inteligentná zmluva je použiteľná na vytvorenie Blockchain Contractual Routing protokolu (BCR) pre zariadenia vo viacpreskokových sieťach. Využíva sa na zistenie cesty zo zdrojového uzla do cieľa (namiesto pôvodného RREQ v typických smerovacích protokoloch) [6], [8]. Vo viacpreskokových sieťach môžu byť uzly v sieti rozdelené do troch častí podľa prenosu dátových paketov [6]:

- **Zdrojový uzol**, ktorý je zodpovedný za prenos dátového paketu.
- **Smerovací uzol**, ktorý je zodpovedný za preposielanie dátového paketu.
- **Cieľový uzol**, ktorý je zodpovedný za príjem dátových paketov.

Smerovací protokol BAODV (Blockchain Ad-hoc On-Demand Distance Vector - BAODV) aplikuje koncepciu reťazenia pomocou hodnoty hašovacej funkcie v BC. Táto hodnota je implementovaná do hlavičky smerovacieho protokolu AODV, ktorá je využívaná v procese smerovania (Route Request packet - RREQ) a overovania uzlov vo vnútri objavenej cesty (Route Replay packet - RREP). Na tento účel sa k RREQ a RREP pridávajú ďalšie polia v hlavičke voľby smerovania. Postup smerovania je mierne upravený tak, aby zahŕňal požadované vylepšenia. RREQ do hlavičky pridá 32-bitové pole, ktoré obsahuje IP adresu uzla a tento uzol musí byť vylúčený z procesu objavovania. RREP do správy pridá dve 256-bitové polia na prenos aktuálnych a aktualizovaných hodnôt hašovacej funkcie blokov a 32-bitové pole, ktoré je pridané na uchovávanie škodlivej IP adresy [1], [9].


```
42.502 Node ID: 45 MAC: 00:00:00:00:00:2e
Source IP: 10.1.1.13
Destination IP: 10.1.1.46
Previous Hash: 6e694f81a1c773b76d84cad4ec524b6307d55df4c3d757abf7aa0704c6e77900
Current Hash: 3d7b18e78453ade6ea80773e20b5168f15338dcc87a118a999b5b9e8171a6bf0
Data: 44
Route: 12 28 44 5 45 Route length: 5
```

Obr. 3: MANET a SHA256

Parameter	Hodnota
Počet uzlov	50 a 100
Simulačná plocha [m ²]	1500 x 1500
Vysielač výkon [dBm]	1 - 7.5
Rýchlosť pohybu [m/s]	1 - 5
Model pohybu	Náhodný model pohybu
Počet opakovaní	5

Tabuľka 1: Parametre simulácie

Počas každej transakcie sa inicializoval nový reťazec a predchádzajúca doba prevádzky reťazca bola ukončená na overenie každej transakcie v sieti, aby bolo zabránené škodlivému konaniu pri každej transakcii. Ak bol objavený škodlivý uzol v sieti počas transakcie, protokol BAODV odovzdá adresu škodlivého uzla zdroju vždy, keď je uzol detegovaný nasledujúcim uzlom, aby inicioval novú správu RREQ s adresou škodlivého uzla vloženou do smerovacieho uzla. Samotná idea fungovania daného protokolu je znázornená na obr. 2.

II. SIMULÁCIA A VÝSLEDKY

Daný smerovací protokol BAODV bol implementovaný do simulátora NS-3 [10]. NS-3 simulátor poskytuje modely fungovania a výkonu paketových dátových sietí a poskytuje nástroj na vykonávanie simulačných experimentov. NS-3 simulátor umožňuje simulácie, ktoré je ťažšie alebo nemožné vykonať so skutočnými systémami, študovanie správania systémov v kontrolovanom a reprodukovateľnom prostredí a získavanie informácií o fungovaní sietí. NS-3 sa zameriava na modelovanie fungovania internetových protokolov a sietí.

Ako komunikačné rozhranie je využitý štandard IEEE 802.11b (WiFi). Čas simulácie bol obmedzený z hardvérových dôvodov na 600 sekúnd. Samotné simulácie boli vykonané na MANET sieti, pozostávajúcej z 50 a 100 mobilných uzlov náhodne rozmiestnených na danej simulačnej oblasti. Samotné uzly sa pohybujú náhodným modelom pohybu. Podrobnejšie nastavenie simulácii je uvedené v tabuľke 1. Hodnoty rýchlosti pohybu uzlov v sieti a ich počet je možné zmeniť. Taktiež je možné meniť aj vysielač výkon siete a aj veľkosť poľa, v ktorom sa nachádzajú uzly. Do simulácie je pridaná hašovacia funkcia SHA-256, ktorú využíva BC (Obr. 3). V programe je zahrnutý aj výstup zo simulácie v podobe .csv súboru, ktorý obsahuje všetky dôležité parametre potrebné pre vytvorenie výsledkov. Jednotlivé parametre sa menia priamo v súbore *manet-blockchain.cc*, ktorý slúži na analýzu a testovanie činnosti AODV protokolu s implementovanou BC. Zdrojový uzol a cieľový uzol sa mení priamo v termináli, cez ktorý sa spúšťa aj samotná simulácia. *ENTRY_NODE_ID* parameter slúži na výber zdrojového uzla a *DEST_NODE_ID* parameter slúži na výber cieľového uzla. Ak je príkaz *.waf -run manet-blockchain.cc -vis* spustený bez parametrov, automaticky sa vyberú uzly 0 a 1 ako zdrojový a cieľový uzol. Prepínač *-vis* slúži na zobrazenie simulácie. Na Obr.4 je ukážka 50 a 100 uzlovej siete, táto topológia sa mení zakaždým, keď sa spustí nová simulácia. V simulácii je možnosť aj presunúť kurzor nad jeden z modrých uzlov a zobrazia sa viaceré informácie o uzle ako jeho číslo v sieti, IP adresa, MAC adresa.

Cieľom simulácii bolo overenie si možností implementácie technológie BC do prostredia simulátora NS-3 a skúmanie, ako daný algoritmus ovplyvní sieťové vlastnosti smerovacieho protokolu BAODV. Boli analyzované parametre *priemerný počet preskokov* a *priemerný počet blokov*.

Parameter priemerný počet preskokov signalizuje priemerný počet preskokov medzi zdrojovým a cieľovým uzlom. Dosiahnuté výsledky sú znázornené v tab. 2a. Z dosiahnutých výsledkov môžeme konštatovať, že BC nezvýši priemerný počet preskokov v danej sieti a samotná implementácia BC do procesu smerovania poskytne možnosti výberu odolnej komunikačnej cesty.

Parameter priemerný počet blokov BC analyzuje zahltenie siete MANET vytvorenými BC správami (blokmi). Tento parameter ovplyvňuje priepustnosť komunikačného kanála samotnými blokmi BC. Dosiahnuté výsledky sú znázornené v tab. 2b. Samotný proces BC generuje veľké množstvo blokov, ktoré ovplyvňujú samotnú kapacitu komunikačného kanála.

III. ZÁVER

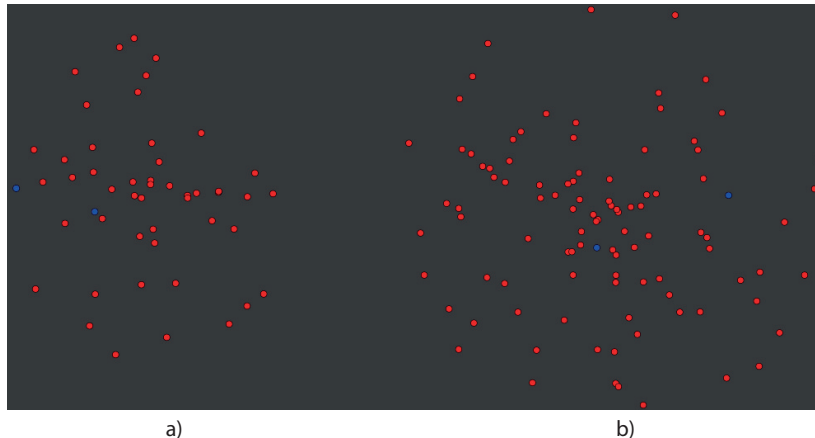
Táto práca sa venuje problematike implementovania technológie Blockchain do prostredia MANET sietí s využitím smerovacieho protokolu AODV (BAODV). V práci sme sa zamerali

Výkon [dBm]	Počet uzlov	Rýchlosť uzlov [m/s]					Výkon [dBm]	Počet uzlov	Rýchlosť uzlov [m/s]				
		1	2	3	4	5			1	2	3	4	5
1	50	4.45	3.81	3.93	2.70	5.98	1	50	607	299	539	352	93
	100	4.58	6.87	4.91	4.28	5.31		100	715	347	428	502	343
2	50	4.83	3.82	3.66	3.47	3.75	2	50	409	426	413	372	284
	100	4.93	3.84	5.01	4.49	5.31		100	730	507	373	561	394
3	50	4.12	4.02	4.86	3.88	3.75	3	50	470	457	552	515	276
	100	5.79	5.82	6.54	7.88	4.86		100	690	516	441	310	299
5	50	4.34	4.42	5.59	5.41	4.26	5	50	822	680	354	398	717
	100	5.13	3.49	4.39	5.46	5.18		100	690	672	510	265	255
7.5	50	4.66	4.03	4.02	5.07	3.89	7.5	50	963	777	800	613	792
	100	4.37	3.04	3.32	3.61	3.20		100	458	846	700	565	627

(a) Priemerný počet preskokov

(b) Priemerný počet BC blokov

Tabuľka 2: Simulovanie technológie BC a MANET



a)

b)

Obr. 4: Topológia MANET v NS-3: a) 50 uzlov, b) 100 uzlov

na samotný proces implementovania BC do prostredia simulátora NS-3. Dosažené výsledky poukazujú na skutočnosť, že daná technológia umožní nájdenie odolnej komunikačnej cesty nielen pomocou smerovacích algoritmov, ale aj pomocou zdieľania si dodatočných informácií uložených v BC medzi jednotlivými uzlami. Dôležitým elementom BC v MANET bude aj možnosť eliminovania nežiadúcich redundantných blokov BC, ktoré v negatívnom zmysle ovplyvnia výkonnosť danej siete.

PODĀKOVANIE

Táto publikácia vznikla vďaka podpore v rámci Operačného programu Integrovaná infraštruktúra pre projekt: Výskum v sieti SANET a možnosti jej ďalšieho využitia a rozvoja, kód ITMS: NFP313010W988, spolufinancovaný zo zdrojov Európskeho fondu regionálneho rozvoja.

LITERATÚRA

- [1] M. A. A. Careem and A. Dutta, "Reputation based routing in manet using blockchain," *COMSNETS*, pp. 1–6, 2020.
- [2] M. Baumgartner and J. Papaj, "Robust data transmission in 5g networks without infrastructure based on blockchain technology," in *2022 32nd International Conference Radioelektronika (RADIOELEKTRONIKA)*, 2022, pp. 01–04.
- [3] A. Jarjis and G. Kadir, "Blockchain authentication for aodv routing protocol," in *2020 Second International Conference on Blockchain Computing and Applications (BCCA)*, 2020, pp. 78–85.
- [4] G. Liu, H. Dong, and Z. Yan, "B4sdc: A blockchain system for security data collection in manets," in *ICC 2020 - 2020 IEEE International Conference on Communications (ICC)*, 2020, pp. 1–6.
- [5] G. Liu, H. Dong, Z. Yan, X. Zhou, and S. Shimizu, "B4sdc: A blockchain system for security data collection in manets," *IEEE Transactions on Big Data*, vol. 8, no. 3, pp. 739–752, 2022.
- [6] A. Yazdinejad, R. M. Parizi, A. Dehghantanha, Q. Zhang, and K.-K. R. Choo, "An energy-efficient sdn controller architecture for iot networks with blockchain-based security," *IEEE Transactions on Services Computing*, vol. 13, no. 4, pp. 625–638, 2020.
- [7] D.-G. Zhang, P.-Z. Zhao, Y.-y. Cui, L. Chen, T. Zhang, and H. Wu, "A new method of mobile ad hoc network routing based on greed forwarding improvement strategy," *IEEE Access*, vol. 7, pp. 158 514–158 524, 2019.
- [8] A. Anand, H. Aggarwal, and R. Rani, "Partially distributed dynamic model for secure and reliable routing in mobile ad hoc networks," *Journal of Communications and Networks*, vol. 18, no. 6, pp. 938–947, 2016.
- [9] R. Bruzgiene, L. Narbutaite, and T. Adomkus, "MANET Network in Internet of Things System. Ad Hoc Networks," *Ad Hoc Networks*, vol. 39, pp. 187–220, June 2017.
- [10] P. D. Bugarcic, M. Z. Malnar, and N. J. Jevtic, "Modifications of aodv protocol for vanets: performance analysis in ns-3 simulator," in *2019 27th Telecommunications Forum (TELFOR)*, 2019, pp. 1–4.

Výber relevantných atribútov asociálneho chovania

¹Kristína MACHOVÁ, ²Michaela HORVÁTHOVÁ

^{1,2} Katedra kybernetiky a umelej inteligencie, Fakulta elektrotechniky a informatiky, Technická univerzita Košice

¹kristina.machova@tuke.sk, ²michaela.horvathova@student.tuke.sk

Abstrakt — Článok sa venuje problému výberu atribútov typických pre asociálne chovanie. Obsahuje popis nástrojov šírenia a odstraňovania príspevkov antisociálneho správania, procesov detekcie falošných správ a spôsobov ako s týmto problémom bojovať. Práca sa zameriava na výber relevantných atribútov a vytvorenie dátovej množiny. V práci je použitá metóda náhodných stromov na generovanie modelu pre rozpoznávanie antisociálneho správania. Na záver je model testovaný a vyhodnocuje sa miera relevancie zadefinovaných atribútov.

Kľúčové slová — online diskusia, antisociálne chovanie, relevancia atribútov, strojové učenie, náhodné lesy, rozhodovacie stromy

I. ÚVOD

V 21. storočí sa dynamicky rozšíril dosah falošných správ na webe spolu s využívaním tohto výrazu. Otvorenie webového prostredia pre ľudí v deväťdesiatych rokoch bolo značným posunom umožňujúci prístup k dátam. Postupom času sa internet rozrástol do nepredstaviteľných foriem s veľkým počtom informácií, ktoré sa neprestajne množia. Práve pre toto sa stal internet hositeľom viacerých nepotrebných, nereálnych a klamlivých informácií, ktoré môže zdieľať a zverejňovať hocikto. Informácie sú v dnešnej spoločnosti dostupné pre širokú verejnosť jedným klikom myši. No nie všetky informácie priniesli pravdivé, presné, objektívne a nestranné zdroje. Aj napriek ľahkému prístupu k vedomostiam v súčasnosti ľudia žijú v dobe definovanej dezinformáciami. Falošné správy, ktoré sa šíria na najznámejších platformách, môžu polarizovať verejnú diskusiu, vytvárať napätie a manipulovať. Už dnes existuje množstvo detektorov antisociálneho správania, či rozšírení prehliadača, ktoré upozorňujú nato, či je webová stránka alebo profil sociálnej siete známy tým, že produkuje zavádzajúce správy.

II. DETEKCIA ANTISOCIÁLNEHO PRISPIEVANIA V ONLINE PRIESTORE

Detekcia antisociálneho prispievania na webe je novo vznikajúcou oblasťou výskumu, ktorá reprezentuje nové myšlienky, priťahuje výskumníkov a má špecifický charakter. Falošné správy sú účelovo napísané pre oklamanie čitateľov, aby týmto nereálnym informáciám uverili. Preto je nutné zahrnúť pomocné údaje, akými sú povedzme spoločenské vzťahy používateľov na sociálnych médiách pre nájdenie znakov – atribútov typických pre falošné správy. Používanie daných pomocných informácií nie je nijako triviálne.

Samotné falošné správy nie sú novým problémom. Ekológia médií falošných správ sa postupom času mení, z novín na rozhlas a z televízie na online správy a sociálne médiá. Falošné správy buď na webe alebo sociálnych sieťach majú vlastné špecifické znaky. Povedzme škodlivé účty, ktoré sú dosť intenzívne a jednoducho vytvorené na zvýšenie šírenia klamlivých správ, ako sú sociálne roboty, či trolovia. Práve z danej príčiny majú používatelia snahu vytvárať skupiny, v ktorých sa združujú totožne zmýšľajúci používatelia, kde polarizujú vlastné postoje [1].

Spôsobom akým ľudia tieto pojmy používajú, vnímajú a rozmyšľajú o nich, sa zásadne rokmi menia. Detekcia falošných správ v sociálnych médiách predstavuje nové a náročné výskumné problémy. Nárast webových správ na sociálnych weboch dáva falošným správam väčšiu silu. Tento problém má niekoľko charakteristík, vďaka ktorým je automatická detekcia náročná. Po prvé falošné správy sú úmyselne písané klamlivým čitateľom, čo ich robí netransparentnými na základe obsahu správ [2].

Práca [3] skúma použitie techník spracovania prirodzeného jazyka na zisťovanie „falošných správ“, to znamená zavádzajúcich správ, ktoré pochádzajú z podozrivých zdrojov. Na základe súboru údajov získaných zo signálu Signal Media a zoznamu zdrojov z OpenSources.com bolo spracovaných asi 11 000 článkov pomocou metódy TF-IDF bi-gramov a pravdepodobnostnej kontextovej gramatiky (PCFG). Testoval sa dataset na viacerých klasifikačných algoritmoch - Support Vector Machines, Stochastic Gradient Descent, Boosting Gradient, Bounded Decision Trees a Random Forests. V práci sa uvádza, že dáta spracované TF-IDF bi-gramov a použité na učenie modelu stochastického gradientu identifikuje nedôveryhodné zdroje s presnosťou 77,2%.

Reklamná platforma Facebooku dáva možnosť tvorcom klamlivých správ, aby vlastné dezinformácie rozširovali s účelom nárastu zisku práve vďaka zobrazovaniu reklám za pomoci daných klamlivých článkov. V následku toho sa firma Facebook pokúsila aplikovať prostriedky na detekciu a boj proti klamlivým správam. Primárnym riešením je, že používatelia majú príležitosť označiť položku, ktorú chápu ako klamlivú. Na zistenie zdroja sú vytvorené značky, ktoré označujú klamstvo a dáva tak možnosť používateľom získať viac dát o príbehu. Keď primerané množstvo používateľov označí príbeh za klamlivý, znižuje sa frekvencia zdieľaného článku [4]. Spoločnosť Facebook aplikuje umelú inteligenciu na identifikáciu klamlivých príspevkov. Tá rozpoznáva extrémizmus, agresivitu, nenávisť, vyhrážky a ďalšie podoby nevhodných príspevkov [5].

Existuje viacero prostriedkov detekcie a odstraňovania klamlivých príspevkov [6]:

- *Spike* – umožňuje zistiť a predpovedať rozprávkové príbehy, ako aj vírusové príbehy. Analyzuje veľké množstvo informácií zo sveta správ.
- *Hoaxy* – pomáha používateľom zistiť klamlivé spravodajské stránky.
- *Snopes* – internetová stránka pomáhajúca rozpoznať klamlivé príbehy.
- *CrowdTangle* – prostriedok umožňujúci pozorovanie obsahu a včasné odhalenie spoločensky neprijateľného obsahu.
- *Check* – pomáha overiť online porušenie správ.
- *Google Trends* – zmysel správ dokazuje kontrolou vyhľadávania.
- *Le Decodex* – databáza obsahujúca internetové stránky označené ako fake alebo real, podľa reálnosti daných stránok.
- *Pheme* – firma, ktorá urobila technologický skok. Dokáže prečítať realnosť obsahu vytvoreného používateľom a online obsahu.

III. NÁVRH PRÍSTUPU K DETEKCI ANTISOCIÁLNEHO OBSAHU

A. Použité metódy

Rozhodovacie stromy, anglicky Decision trees, sú jedným z najstarších a najpoužívanejších algoritmov, ktoré sa využívajú na klasifikáciu a predikciu. Sú priehľadné a jednoducho interpretovateľné, čiže používateľom poskytuje rýchle a prehľadné vyhodnotenie dosiahnutých výsledkov. Cieľom rozhodovacích stromov je na základe prvkov z tréningovej množiny, ktoré sú vstupnými údajmi, predpovedať hodnotu cieľovej premennej. Jeho štruktúra je grafická vo forme stromu, konštruovaná spôsobom zhora-nadol, ktorý sa skladá z hrán a uzlov. Uzly sú triedami alebo testovacím atribútom a hrany sú hodnotami testovacieho atribútu. Pre výber testovacieho atribútu sa využíva informačná entropia, tzv. informačný zisk [7].

Algoritmus náhodný les, anglicky Random forest, je používaný pre klasifikáciu aj regresiu. Realizuje takzvané učenie súborom metód. Ide o pomerne novú metódu, ktorá nevyžaduje pred učením modelu zmenšenie priestoru atribútov. Princípom je generovanie väčšieho počtu rozhodovacích stromov, ktoré sú navzájom dekorované a následne sa formuje finálne rozhodnutie o klasifikácii hlasovaním alebo spriemerovaním. Tiež hovoríme, že sa jedná o nadstavbu rozhodovacích stromov, odstraňujú ich viaceré nedostatky ako napríklad nestabilitu. Pri veľmi veľkom množstve rozhodovacích stromov, ktoré nie sú upravené orezaním fungujú efektívne na tréningovej množine, no často zlyhávajú na testovacej množine. Tento problém rieši taktiež náhodný les. Algoritmus náhodný les je tvorený z viacerých rozhodovacích stromov CART alebo C4.5, kde je dostupná len náhodná podmnožina atribútov, ktorá je nezávislá od ich počtu. Náhodný les je závislý od sily jednotlivých klasifikátorov a od miery závislosti medzi nimi [1].

B. Popis dát

Dáta použité v tejto práci, sú extrahované z rôznych dôveryhodných a nedôveryhodných portálov (ako napr. sme.sk, dennik.sk či hlavnespravy.sk) a použité na formovanie dátovej množiny, ktorá obsahovala príspevky spadajúce pod antisociálne správanie ako aj neutrálne príspevky. Dataset, ktorý obsahoval zozbierané príspevky bol prístupný vo formáte .xlsx (Excel). Vyskytovalo sa v ňom približne 200 príspevkov. Príspevky boli rôzneho typu antisociálneho

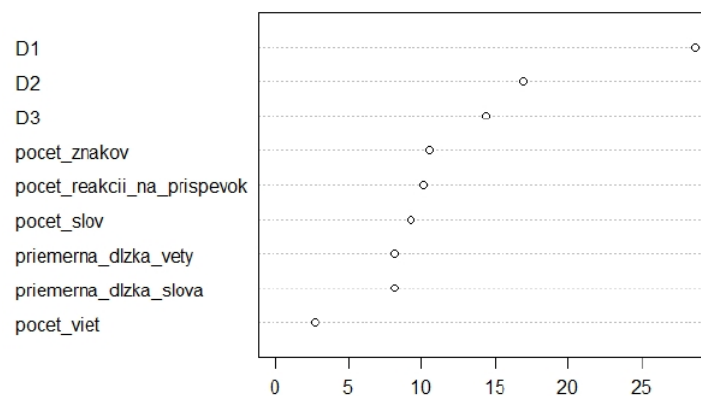
správania a to konkrétne neutrálne príspevky označené s číslicou 0, útočné príspevky označené číslicou 1, negatívne príspevky označené číslicou 2 a manipulačné príspevky, ktoré sme označili číslicou 3.

Pri výbere a definovaní atribútov sme sa zamerali hlavne na ich relevantnosť. V datasete sme zadefinovali atribúty ako počet slov, znakov a viet jednotlivých príspevkov. Ďalšími atribútmi bola priemerná dĺžka slova, ktorú sme vypočítali ako počet znakov delené počet slov, ďalej atribút priemerná dĺžka vety, ktorú sme počítali ako počet slov delené počet viet. Tieto dva atribúty sme použili hlavne nato, aby sme zistili aké dlhé slová a vety používajú v antisociálnych príspevkoch. Následne bol zvolený atribút počet reakcií na článok. Ďalšie atribúty, ktoré sme použili boli tri slovníky špeciálnych slov: D1 (útočný), D2 (negatívny), a D3 (manipulačný) ako slovníky antisociálneho správania, kde jednotlivé slová majú priradené hodnoty, ktoré vyjadrujú váhu slov. Túto váhu slova sme vypočítali a vyhodnotili cez jednoduchý program v programovacom jazyku Java. Posledným zadefinovaným atribútom v dátovej množine je labelovanie. Tento atribút obsahuje hodnoty od 0-3 a to neutrálny 0, útočný 1, negatívny 2 a manipulačný príspevok 3. Hodnoty boli objektívne priradené na základe textu daného príspevku, nie podľa D1, D2, D3.

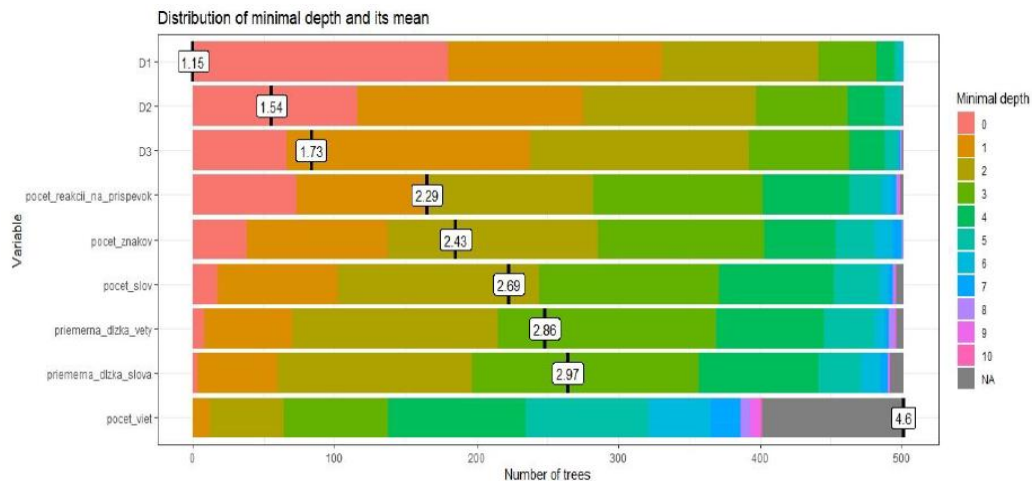
C. Trénovanie a testovanie modelov

Na trénovanie modelov bol použitý algoritmus generovanie lesa náhodných rozhodovacích stromov v prostredí „studia R“. Dáta boli rozdelené na trénovaciu a testovaciu časť v pomere 70% ku 30%. Súbor metód obsahoval 501 náhodne generovaných rozhodovacích stromov.

V prvom rade sme chceli skúmať mieru dôležitosti jednotlivých atribútov pri generovaní jednotlivých stromov, ktorá sa odvíjala od početnosti použitia toho ktorého atribútu pri generovaní stromov. Dôležitosť jednotlivých atribútov je ilustrovaná na Obr.1. Distribúcia atribútov v uzloch stromov vo forme testovacích atribútov je uvedená v Obr.2.



Obr. 1 Miera dôležitosti jednotlivých atribútov viazaná na početnosť výberu pri generovaní rozhodovacích stromov.



Obr. 2 Graf distribúcií atribútov v uzloch stromov.

Na Obr.2 je graficky znázornená distribúcia testovacích atribútov v rozhodovacích stromoch. Atribút D1 je prvý, potom D2 atď. Hodnota D1(1.15) vyjadruje priemer hĺbky atribútu v strome, čiže väčšinou sa nachádza buď prvý alebo druhý, no z grafu vidíme, že v niektorých prípadoch sa používa aj ako piaty.

Výsledky testov modelov naučených metódou Náhodný les rozhodovacích stromov pre 4 triedy (neutrálne, útočné, negatívne a manipulačné príspevky) vo forme hodnôt troch mierok efektivity (Presnosť, Návratnosť a F1) sú uvedené v Tabuľke 1.

Tabuľka 1

Výsledky testov modelov generovaných Náhodným lesom (Random Forest) v hodnotách Presnosti, Návratnosti a F1.

	Neutrálne	Útočné	Negatívne	Manipulačné
Presnosť	0,50	0,95	0,77	0,69
Návratnosť	0,50	0,75	0,87	0,90
F1 Rate	0,50	0,83	0,81	0,78

Podľa hodnôt z Tabuľky 1. sa útočné príspevky dajú najefektívnejšie identifikovať, preto dosahujú najväčšiu presnosť. Pri neutrálnych príspevkoch je klasifikácia ťažšia, pretože ak použijeme slovo z niektorého slovníka aj keď príspevok nie je v nesprávnom kontexte, (napr. slovo „somár“ použité v príspevku o Zoo) tak sa priradí váha k určitej triede a tým pádom nie k neutrálnym.

IV. ZÁVER

Článok definuje atribúty antisociálneho správania a vyčísluje mieru ich dôležitosti pri generovaní lesa rozhodovacích stromov. Taktiež boli testované modely založené na náhodných lesoch a dosiahnuté výsledky sú sľubné.

V budúcnosti by bolo možné na množine najdôležitejších atribútov vyselektovaných v tejto práci naučiť nové modely založené na iných metódach strojového učenia.

POĎAKOVANIE

Práca prezentovaná v tomto článku bola podporená Slovenskou agentúrou na podporu vedy a výskumu v rámci projektu VEGA 1/0685/21.

LITERATÚRA

- [1] HAN, J., KAMBER, M. *Data Mining: Concepts and Techniques*: Morgan Kaufmann Publisher: San Francisco, 2003. 740 s. ISBN 1-55860-901-6 7.
- [2] SHU, K. et al. Fake news detection on social media: A data mining perspective. *ACM SIGKDD Explorations Newsletter*, 2017, 19.1: 22-36.
- [3] AHMED, H., TRAORE, I., SAAD, S. Detekcia falošných správ online pomocou analýzy N-gramu a metód strojového učenia. In: *Medzinárodná konferencia o inteligentných, bezpečných a spoľahlivých systémoch v distribuovaných a cloudových prostrediach*. Springer, Cham, 2017. s. 127-138.
- [4] ŠNÍDL, V.: *Pravda a lož na Facebooku*. Bratislava : N Press, 2017. 157 s. ISBN 978-80- 9723-944-2.
- [5] BYSTRICKÝ, V., ROGUĽOVÁ, J.: *Storočie propagandy*. Bratislava : AEPRESS, 2005. 234 s. ISBN 978-80-8888-064-6.
- [6] OLSON, D., DELEN, D. *Advanced Data Mining Techniques*, ISBN: 978-3-540-76916-3.
- [7] HAN, J., KAMBER, M. *Data Mining: Concepts and Techniques*: Morgan Kaufmann Publisher: San Francisco, 2003. 740 s. ISBN 1-55860-901-6 7.

Identifikácia trollov na základe podobnosti textov

¹Kristína MACHOVÁ, ²Samuel ŠIROKÝ

^{1,2} Katedra kybernetiky a umelej inteligencie, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach

¹kristina.machova@tuke.sk, ²samuel.siroky@student.tuke.sk

Abstrakt — Článok sa zaoberá identifikáciou trollov v online komunitách na základe podobnosti textov. V krátkosti ponúka prehľad danej problematiky a motiváciu k výskumu v oblasti identifikácie trollov. Príspevok sa sústreďuje na identifikáciu autorov príspevkov už známych trollov. Cieľom je navrhnúť prístup k detekcii trollov založený na podobnosti autorského štýlu. Tento prístup viedol ku generovaniu detekčného modelu pomocou Naivného Bayes klasifikátora, ktorý v spolupráci s metódou TF-IDF použitou pre vektorizáciu textu dosiahol sľubné výsledky.

Kľúčové slová — sociálne médiá, trolizmus, identifikácia trollov, podobnosť textu, autorský štýl, Naivný Bayes klasifikátor, TF-IDF

I. ÚVOD

V internetovom slangu je trol osoba, ktorá začína hádky, snaží sa uraziť ľudí, rozptýliť a zasiať rozkol uverejňovaním poburujúcich, odbočujúcich, alebo nesúvisiacich príspevkov v online konverzáciách, vo fórach, komunitách a podobne. Úmyslom trola je podľa [1] vyprovokovať ostatných používateľov k emotívnym reakciám a odbočeniu od témy konverzácie, či už za účelom vlastného obohatenia, alebo len pobavenia.

Tom Postmes, profesor sociálnej a organizačnej psychológie pôsobiaci na univerzitách v Exeteri v Anglicku a v Groningene v Holandsku hovorí: „Trolovia sa usilujú o násilie na úrovni problémov, ktoré môžu spôsobiť v prostredí. Chcú, aby sa to rozbehlo. Chcú podporovať antipatické pocity hnevu a pobúrenia, ktoré im morbidne dodávajú pocit potešenia“.

Terajšia moderná doba umožňuje takýmto diskutérom pod ochranou práva vyjadrenia vlastného názoru vyjadriť sa ku každej téme, bez ohľadu na ich kvalifikáciu. Avšak narastá tak hrozba manipulácie, hrozba tvorby dezinformácií, hrozba šírenia poplašných správ, či iné ohrozenia podobného charakteru. Spoločnosť pociťuje, že je nutné takéto správanie kontrolovať a dožaduje sa riešenia. Niektorí giganti sociálnych médií už začali konať a pod nátlakom spoločnosti sú nútení ponúkať riešenia tejto situácie. Na základe vyššie spomenutých faktov je našim cieľom navrhnúť prístup k identifikácii už existujúcich trollov na základe štýlu autora príspevku na online fóre, komunikácií, alebo v komunite.

II. IDENTIFIKÁCIA TROLOV

Oxfordský slovník definuje trola ako „osobu, ktorá úmyselne útočí, alebo provokuje v online príspevkoch.“ Dalo by sa povedať, že táto definícia je podhodnotená. A to z toho dôvodu, že v dnešnej dobe správne cieleným trolizmom je možné dosiahnuť oveľa viac, ako len niekoho uraziť. Týmto spôsobom sa dá vyslať falošná správa, ktorá môže spôsobiť masové šialenstvo. Taktiež je možné kariérne zničiť osobu, alebo aj celú spoločnosť. V neposlednom rade sa dajú ovplyvniť aj výsledky volieb. V minulosti by k takýmto úkonom viedli dlhé hodiny plánovania, stratégie a museli by byť využité nemalé finančné prostriedky. Avšak podľa zdroja [2] to už v dnešnej online dobe nie je potrebné. Z týchto dôvodov je veľmi žiadúca automatická identifikácia trollov a obmedzenie ich pôsobenia.

Trollov nie je možné presne určiť pre ich komplikované a nestabilné správanie. Ako sme už spomínali, aj na prvý pohľad neškodné príspevky môžu byť v určitom kontexte veľmi nevhodné a poburujúce. Vyhľadávaním hrubých a vulgárnych výrazov, urážajúcich pomenovaní a nadávok sa dá zachytiť značné množstvo trollov. Ale, čo s tými, čo sú tak povediac „páni slova“? Zatiaľ sa týmito „vyššími trolmi“ zaoberajú ľudskí znalci. Identifikácie trollov za pomoci kontextu ich príspevkov je veľkou výzvou do budúcnosti.

Identifikácia trollov je ako vedná disciplína relatívne nová. V minulosti boli na túto prácu zamestnávaní ľudia, prechádzali nahlásené príspevky a posudzovali ich. Avšak onedlho na to ľudskí moderátori nebudú stačiť. S príchodom trol-botov, čo sú programy vytvorené za účelom trolizmu, ľudské zdroje začínajú zaostávať. Jeden trol-bot by dokázal za hodinu vytvoriť dostatočný počet príspevkov aby zamestnal celý tím znalcov na pár dní. V tejto práci sme sa zamerali na ľudských trollov. Ani v tomto smere by pravdepodobne moderátori nestíhali. Akonáhle ide o väčšie fórum, sieť, komunitu, manuálne odhaľovanie trollov by si už vyžadovalo celý tím špecialistov.

Základnými znakmi trollov podľa [3] formulovanými na účely ich odhaľovania sú:

- *Anonymita* – často požívajú prezývky, pod ktorými môže vystupovať ktokoľvek.
- *Vykonštruované e-maily* – ak stránka, alebo fórum vyžaduje overenie e-mailom, tak si tento e-mail vytvorí len pre tieto účely.
- *Proxy servery* – využívajú proxy servery, aby nebolo možné vystopovať ich IP adresu.
- *Profil* – žiadne, minimum alebo klamlivé informácie v profile používateľa môžu naznačovať, že sa jedná o trola.
- *Fotografia a meno* – ak používateľ nemá fotografiu, je to trochu podozrivé, taktiež ak je to fotografia s vodoznakom po celej ploche. Nick používateľa by nemali byť nejaké náhodné písmená a číslice, ako napríklad: Xhasdqruk, Jozef67230752.
- *Aktivita konta* - ako často a kedy naposledy sa prihlásil, pridáva príspevky pravidelne, alebo raz za čas. Aj správanie používateľa veľa napovie.

III. VERIFIKÁCIA AUTORSTVA

Automatická verifikácia autorov predstavuje prístup, keď meraním určitých parametrov môžeme rozlišovať rôznych autorov, teda či bol daný dokument skutočne napísaný konkrétnym autorom. Dnes rastú online texty, ako boli fóra, blogy, e-maily, chaty, a iné nevídaným tempom a vyžaduje sa teda zefektívnenie prístupu k identifikácii autorov zapojením vedeckých oblastí ako strojové učenie, spracovania prirodzeného jazyka a dolovania informácií. Došlo k odhaleniu potenciálu identifikácie autora v rôznych smeroch. Sú veľké snahy určiť autorov teroristických vyhlásení, alebo rôznych výhražných listov. Overuje sa autenticita samovražedných odkazov. Tieto metódy našli značné využitie aj v súdnych sporoch o autorské práva a napomáhajú aj identifikácii počítačových kriminálnikov.

Úloha určenia autora textu podľa [4] pozostáva z priradenia potenciálneho autora s najvyššou mierou pravdepodobnosti, z množiny potenciálnych autorov, k dokumentu alebo k textu, pričom k dispozícii sú príklady dokumentov všetkých potenciálnych autorov. Z aspektu strojového učenia ide o multi-triednu klasifikáciu.

Podľa zdroja [5] sú známe viaceré prístupy ku identifikácii autora textu. Niektoré prístupy spájajú všetky tréningové texty jedného autora do jedného celku a z neho extrahujú autorský štýl daného autora. Tento proces je nazývaný *Profilovanie autora*, kde sa rozdiely medzi textami toho istého autora neberú do úvahy. Druhá cesta je mať viac textových vzoriek od jedného autora pre naučenie detekčného modelu. To znamená, že každý text je prezentovaný ako samostatná inštancia autorského štýlu. Tieto postupy nazývame *Postupy založené na inštanciách*.

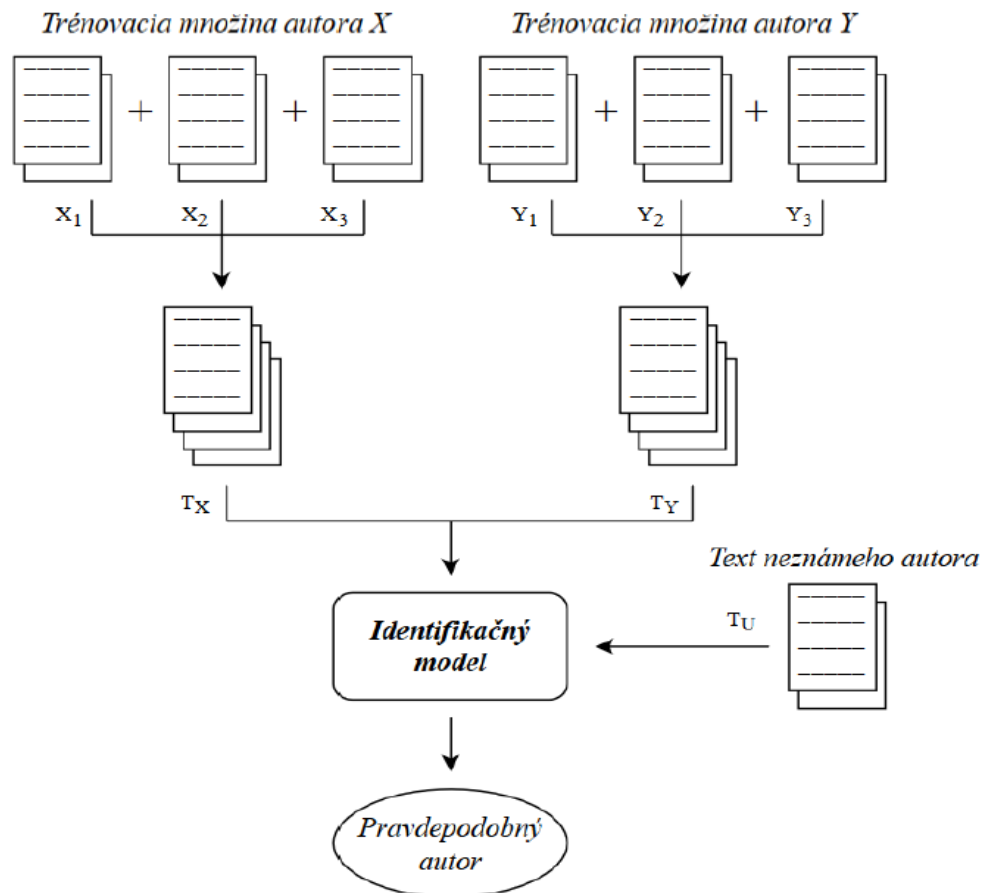
A. Postup založený na profilovaní autora

Profilovacie postupy majú jednoduchú tréningovú fázu, ktorá pozostáva z extrakcie profilov jednotlivých autorov. Klasifikačný model je zvyčajne založený na dištančnej metóde, ktorá počíta rozdiely medzi profilom neznámeho textu a profilom každého autora. Najpravdepodobnejší autor je určený na základe miery vzdialenosti medzi jednotlivými štýlmi. Pri tejto metóde neexistuje žiadna samostatná reprezentácia jednotlivých vzoriek textu. Je iba jedna veľká reprezentačná zložka autora pozostávajúca zo všetkých jeho známych textov. Preto sú rozdiely medzi tréningovými textami toho istého autora indiferentné. Typická architektúra profilových prístupov je znázornená na Obrázku 1.

B. Postup založený na inštanciách

Tento postup je založený na použití metód strojového učenia, ktoré považujú každú tréningovú vzorku textu za samostatnú jednotku, ktorá prispieva k modelu identifikácie. Každá textová vzorka známeho autora je jeho identifikačným príkladom. Klasifikačné algoritmy si vyžadujú viac tréningových príkladov v jednej triede (jedného autora) pre naučenie spoľahlivého klasifikačného modelu. Pri použití tohto prístupu na klasifikáciu autora z jeho jedného textu (napr. kniha) je potrebné tento text rozdeliť na viac častí, ktoré sú rovnako dlhé. Ak existuje viac vzoriek textov s rôznou dĺžkou, je potrebné rozsah textov normalizovať. Tréningové texty musia byť dostatočne dlhé, aby bolo možné zachytiť vlastnosti reprezentácie textu. Uskutočnené

experimenty s rôzne dlhými literárnymi blokmi textov (200, 500, 1000 slov) ukázali, že znížením dĺžky textových blokov dochádza k výraznému zníženiu presnosti. Proces výberu vzorky textu priamo ovplyvňuje výkonnosť identifikačného modelu. Architektúra inštančných postupov je zobrazená na Obrázku 2.



Obr. 1 Architektúra profilovo založených postupov.

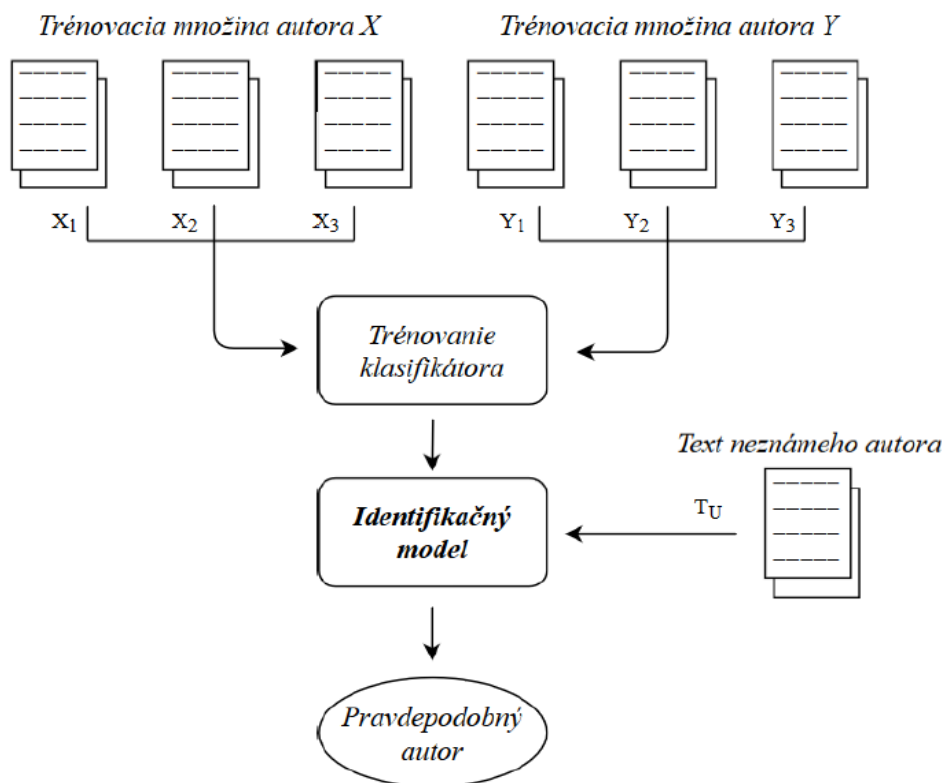
IV. NAVRHNUTÝ PRÍSTUP

Nami navrhnutý prístup bol implementovaný v nasledovných krokoch:

- Načítanie textov
- Vektorizácia textu
- Výber sady atribútov
- Trénovanie modelu
- Testovanie modelov na rozpoznávanie trolov.

Dáta sú do programu načítané zo súboru *Databaza_Final-C.csv*, kde prípona *.csv* označuje typ súboru nazývaný ako comma-separated values. Comma-separated values v doslovnom preklade znamená hodnoty oddelené čiarkami. Kódovanie súboru je typu UTF-8. Súbor sa načíta po riadkoch, ktoré sú rozdelené do dvoch stĺpcov. Následne sú rozdelené do dvoch setov, t.j. do testovacej a trénovacej množiny. Spomínaný dataset sme vytvorili extrahovaním zo zdrojov SME.sk, Pravda.sk a mojeVideo.sk. Dataset obsahoval 500 príspevkov od 100 troľích autorov.

Pri vektorizácii textu sme použili metódu TF-IDF. Je to jedna z najjednoduchších štatistických metód a nie je zložitá ju naprogramovať. V našom riešení sme sa rozhodli použiť funkciu z knižnice Scikit-learn (*sklearn.feature_extraction.text.Tfidfvectorizer*). Daná funkcia obsahuje dva parametre: *sublinear_tf* a *max_df*, ktoré nastavíme nasledovne. *Sublinear_tf* bude rovné hodnote *true* a to zabezpečí, že vo výpočte bude parameter *tf* zmenený na $1 + \log(tf)$. Týmto krokom odstránime problém vysokého výskytu určitého slova. To znamená, že pätnásť výskytov jedného slova nebude predstavovať pätnásťnásobok významu tohto slova. Ak *max_df* nastavíme na hodnotu napríklad 0,6 tak budeme ignorovať výrazy, ktoré majú frekvenciu výskytu v dokumente vyššiu ako 0,6. To znamená, že funkcia nebude brať do úvahy stop words (napríklad a, ak, aby, my, na ...).



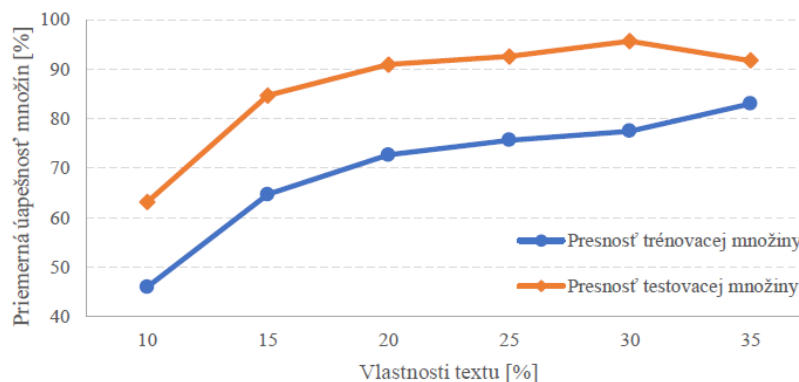
Obr. 2 Architektúra inštančne založených postupov.

Výber sady atribútov textu je vhodný pri objemných textoch, kde by mohlo byť výpočtovo náročné trénovať model s príliš veľkým počtom slov – priveľkým lexikálnym profilom. Aj v tomto prípade sme použili už vytvorenú knižnicu *SelectPercentile* zo Scikit-learn pre výber atribútov textu s najvyšším skóre. *SelectPercentile* má dva parametre. Prvý parameter je skórovacia funkcia a druhým parametrom je percentil skórovacej funkcie. Ako skórovaciu funkciu sme zvolili *f_classif*, ktorá počíta F-hodnoty ANOVA pre vzorku. Cieľ bol výber slov, alebo termínov, ktoré dosahujú najvyššie F-hodnoty, takže slová, alebo termíny, ktoré sa s najväčšou pravdepodobnosťou líšia v závislosti od triedy, alebo autora. Ide o bežný postup pre výber najlepších diskriminačných vlastností medzi triedami.

Pri trénovaní modelu sme použili naivný Bayesov klasifikátor. Vo všeobecnosti trénovanie modelu strojového učenia s knižnicami Scikit-learn má nasledovný postup: inicializácia inštancie modelu triedy, prispôbenie trénovacích dát, predpoveď testovacích dát a výpočet presnosti (angl. accuracy) pre testovací a trénovací set. Výstupom modelu sú nasledovné údaje: odhadovaný trénovací čas, odhadovaný testovací čas, presnosť trénovacej množiny a presnosť testovacej množiny.

A. Výsledky testov a ich analýza

Najprv sme skúmali efektívnosť detekčných modelov v závislosti na veľkosti množín atribútov, respektíve vlastností použitých na trénovanie modelu. Výsledky sú uvedené v Obrázku 3.



Obr. 3 Presnosť modelov na detekciu trolov v závislosti od mohutnosti množiny vlastností – atribútov použitých na trénovanie modelu NB klasifikátora.

Napokon sme uskutočnili experimenty, kde sme nastavili výber sady atribútov na 35% a dataset sme rozdelili na tréningovú a testovaciu časť v pomere 90/10 a uskutočnili desať násobnú krížovú validáciu. Výsledky sú uvedené v Tabuľke 1.

Tabuľka 1

Výsledky testov s desaťnásobnou krížovou validáciou detekčného modelu založeného na Naivnej Bayes metóde.

Pokus	Tréningový čas [s]	Testovací čas [s]	Presnosť tréningovej množiny [%]	Presnosť testovacej množiny [%]
1	0,95485	0,01801	80,49	95,29
2	0,98687	0,01895	78,22	94,12
3	0,74949	0,01396	78,22	95,69
4	0,98585	0,01896	81,60	94,90
5	1,00000	0,03890	80,09	96,47
6	0,94308	0,01596	77,11	94,90
7	0,92777	0,01696	79,51	94,12
8	0,98181	0,01795	82,93	95,69
9	0,96494	0,04016	78,44	94,51
10	0,96964	0,01995	80,27	97,25
Priemer	0,95485	0,02198	80,49	95,29

V. ZÁVER

V článku je uvedený popis prístupu k identifikácii trolov v online komunitách. Kľúčovým faktorom bola podobnosť textov a štýl písania. Výsledkom našej práce je implementácia, ktorá porovnáva používateľom zadaného autora a jeho texty s našim datasetom a výsledkom je vyhodnotenie s akou percentuálnou pravdepodobnosťou sa jedná o trola. Vytvorenie datasetu bolo časovo náročné, pretože sme sa zameriavali na kontextovú stránku príspevkov. Zozbierať sto trolov s ich cca piatimi príspevkami, pre naše experimenty relevantnými, nám trvalo približne 3 mesiace. Tento dataset bol pre našu prácu postačujúca množina, avšak pre budúce experimenty s tréningom modelov by bolo vhodné množinu príspevkov od trolov zväčšiť.

Experimenty naznačujú, že Naivný Bayes klasifikátor spolu s TF-IDF metódou pre vektorizáciu textu a výberom vlastností textu by mohli dosiahnuť v úlohách identifikácie autorov typu trol môže poskytnúť sľubné výsledky. Obsahom našej práce je jedna konkrétna metóda určovania autorstva. V budúcnosti by bolo vhodné rozšíriť tento výskum zapojením ďalších metód strojového učenia. Z uskutočnených pokusov vyplynulo, že optimálnejšie nastavenie je pri rozdelení datasetu v pomere 90/10 tréning/testovanie a nastavení výberu atribútov textu na 35%. Pri takom nastavení náš model dosiahol približne 95% presnosť pri rozpoznávaní trolov. Súčasťou práce je aj aplikácia, ktorá po vložení používateľského vstupu vo forme textu vyhodnotí s akou pravdepodobnosťou je autorom vloženého textu trol.

POĎAKOVANIE

Práca prezentovaná v tomto článku bola podporená Slovenskou agentúrou na podporu vedy a výskumu v rámci projektu VEGA 1/0685/21.

LITERATÚRA

- [1] COLLINS, X. Definition of „trol“ [online]. Collins English Dictionary. Dostupné na internete: <https://www.collinsdictionary.com/dictionary/english/troll>, [cit. 30.4.2022].
- [2] BHATT, U. IYYANI, D. JANI, K. MALI, S. Troll-detection systems: Limitations of troll detection systems and AI/ML anti-trolling solution. 3rd International Conference for Convergence in Technology (I2CT), India, Apr 06-08, 2018.
- [3] Deborah, N.G. How to identify trolls in your online community, Dostupné na internete: <https://www.dummies.com/social-media/blogging/how-to-identify-trolls-in-your-online-community/>, [cit. 30.4.2022].
- [4] STAROŇOVÁ, P. Pripísanie autorstva. Diplomová práca. Vydavateľ: Technická univerzita v Košiciach, 2018, ps. 54.
- [5] EFSTATHIOS, S. A Survey of Modern Authorship Attribution Methods. Journal of the American Society for Information Science and Technology. Volume 60 Issue 3, March 2009, pp. 538-556.

An Inverse Fuzzy Model Strategy for Dynamic System Control

¹Pavol FEDOR, ²Peter Radváni

¹Katedra elektrotechniky a mechatroniky, Fakulta elektrotechniky a informatiky,
Technická univerzita v Košiciach, Slovenská republika

²EnergoControl s.r.o., Pri plynární 2, 040 01 Košice, Slovenská republika

pavol.fedor@student.tuke.sk, radvani@energocontrol.sk

Abstract — The paper deals with the design and simulation of properties of an unknown dynamic system controller based on its inverse fuzzy model. In the first step of the design process, a black-box system fuzzy model is formed on basis measured input-output data. The model is then exploited for creating an inverse fuzzy model of the system. The paper demonstrates one of the possibilities of an inverse fuzzy model application for purpose of control. The method is first explained by means of simple examples, followed by the presentation of results of its application in the control of a drive with asynchronous motor. The design results are validated by simulation in the MATLAB and by experimental measurements on an AC inverter–asynchronous motor system, carried out by means of a Real Time system.

Keywords — nonlinear dynamic system, inverse fuzzy model, black-box system, asynchronous motor

I. INTRODUCTION

Controlling systems with an unknown analytical description and unknown parameters presents nowadays a very demanding task. The problem issue is obtaining the description of these qualitative properties on basis of experimental data, which, without prior knowledge of the parameters and structure of the system being described, usually results in a non-consistent database, in problems with covering the entire possible space of fuzzy inputs into the system, etc. The fuzzy model thus obtained is then often practically unusable, although if a different way of data acquisition is chosen, or if suitable choice of the „correct“ data from the database is made. The essence of this problem is obtaining information about the approximate qualitative properties of the system controlled by means of its fuzzy model formed on basis of appropriate measurements and the subsequent exploitation of this information for the design of the appropriate (very often also fuzzy) controller [1]- [3]. One possible example of this procedure is demonstrated in the following paper.

II. PROBLEM DESCRIPTION

Let us assume we are looking for a simplified fuzzy description of a nonlinear dynamic system with a single input u and single output y , described as:

$$\begin{aligned} dx &= \mathbf{H}(u_m, \mathbf{x}) \\ y &= x_1 \\ \mathbf{H} &= [h_1, h_2, \dots, h_n] \end{aligned} \quad (1)$$

where function $\mathbf{H} = [h_1, h_2, \dots, h_n]^T$ is an unknown nonlinear continuous function, \mathbf{x} is a state vector and n is the system order. Let us assume that the monitored and measured output variable of the system is the first state variable, i.e. $y = x_1$. We want to substitute this system by a simplified dynamic first-order system in the form:

$$\begin{aligned} dx &= \mathbf{f}(u_m, \mathbf{x}) \\ y &= x_1 \end{aligned} \quad (2)$$

in which $f(u, x)$ is the unknown nonlinear function, obtained from suitable measurements in system (1). The criterion for substitution is the best possible similarity of responses of systems (1) and (2) for identical input variables u .

We are further looking for such function $g(u, x)$ that fulfils the following condition:

$$\dot{y} = f(g(u, x), x) = K \cdot u \quad (3)$$

By the application of the above procedure we can obtain from the nonlinear system (1) an „approximately“ linear system, the dynamics (derivation) of whose output variable $y \equiv x$ can be directly controlled by input u .

In the literature [4] is shown the method of linearization of nonlinear systems, which are described analytically and it is possible to express functions required for the linearization from their mathematical description. If the system is difficult to describe mathematically or there no exist at all its mathematical description, then another method of the system description have to be chosen. One possible method of system identification, i.e. obtaining quantitative information about the system, is to measure the dependencies between its inputs and outputs.

The algorithm of the procedure may consist of the following steps:

- determining a substitute function f for the whole space of inputs and outputs from a suitably measured database of inputs and outputs of system (1);
- determining a function g (from the prescribed desired equation (3) and function f), which covers the entire work space of the system output and its derivative.

III. THE METHOD OF FUZZY MODEL DETERMINATION

The basic premise for the following identification of system (1) is that it is stable and at constant input u_m its output settles, after fading of the transient performances, at the constant value y . This means that the transition of some measured system (1) from one steady state A to another steady state B at constant input u_m determines one relation among the relevant three values $[u_m, x, dx]$, i.e. one point of function f where $y = x$. The database of measured data for identification of function f should cover the whole work space of pairs $[u_m, x]$, i.e. it is necessary to measure the desired transitions between the steady states in this space. If we split the input space u_m and thus also the relevant space of steady outputs (1) into n parts, then the number of desired measured responses is equal to the number of variations of n elements in groups of 2, i.e.

$$V_2(n) = n!(n - 2)! \quad (4)$$

It should be noted that while the axis of input variable u_m can be evenly divided over the whole range, the axis of output variable y is, in general, divided unevenly, according to the particular system (1).

Example 1: Consider a simple first-order nonlinear dynamic system in the form:

$$\dot{x} = u_m - \sin(x) \quad (5)$$

The working space for the input u_m is defined from the range $\langle 0; 3 \rangle$. Let us split the input space u_m into 5 levels with values $[0; 0.75; 1.5; 2.25; 3]$. In this case it is necessary to measure 20 transitions between steady states and use them to determine identification points for function f . Since this is a first order nonlinear system, the value dx has been determined for each pair of inputs $[u_m, x]$ easily as is shown in Fig. 1.

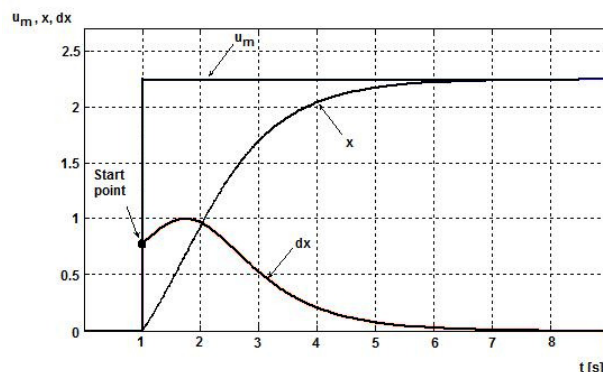


Fig.1 The transient performance detail

The start points represent the peaks for fuzzy description of function (5) in accordance with Table 1.

Table1 Identification points for function f

$U_m \backslash X$	0	0,75	1,5	2,25	3
0	0	-0,6816	-0,9975	-0,7781	-0,1411
0,75	0,6816	0	-0,6816	-0,9975	-0,7781
1,5	0,9975	0,6816	0	-0,6816	-0,9975
2,25	0,7781	0,9975	0,6816	0	-0,6816
3	0,1411	0,7781	0,9975	0,6816	0

A basis points on the diagonal are values of system steady states. This table is for building a fuzzy description (model) of function f obtained by Anfisedit tool of MATLAB. The method used for the fuzzy system generation was the subclustering method, having the following parameters:

Range of influence = 0,5

Squash factor = 1,25

Accept ratio = 0,5

Reject ratio = 0,15.

Function f is represented by Sugeno-type fuzzy system with 9 rules, the characteristic surface of which is shown in Fig. 2.

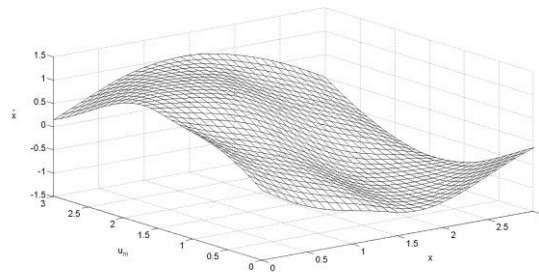


Fig.2 Fuzzy model of identified function f

In this example the first order nonlinear system has been linearized similarly by the first order system. Responses of the original system and of its fuzzy substitute according to (2) to the identified transitions are compared in Fig. 3.

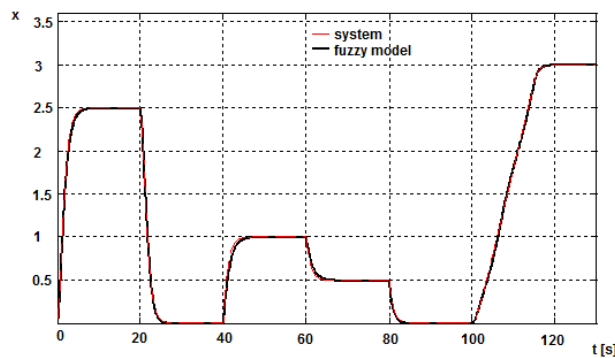


Fig.3 Comparison of the dynamic system and its fuzzy substitute

Because higher-order systems (1) are usually approximated by a first-order system (2), for each particular measured response its best corresponding value dx should be determined. One of the possible suitable procedures is illustrated in Fig. 4, where the beginning of transition of system (1) from one steady point A is substituted by such constant acceleration $dx = \Delta y / \Delta T$ that corresponds to the transition of system (2) to point B determined as the first extreme value of dx following after start of the transition.

IV. METHOD OF FUZZY MODEL INVERSION

The role of function g is to adapt input u_m to the nonlinear system in such a way that the first derivative of the output value corresponds, as much as possible, to the control input u . In the preceding section we illustrated the procedure of obtaining a fuzzy description of function $dx =$

$f(u_m, x)$ on the basis of experimental measurements. From this relation it is then possible to determine for each particular desired value dx and for given system output $y=x$ the corresponding value of u_m . For linearization of system (1) the desired values dx are determined by equation (3), in which the amplification K is usually chosen such that it regulates the range of real values dx to the range of the control signal u .

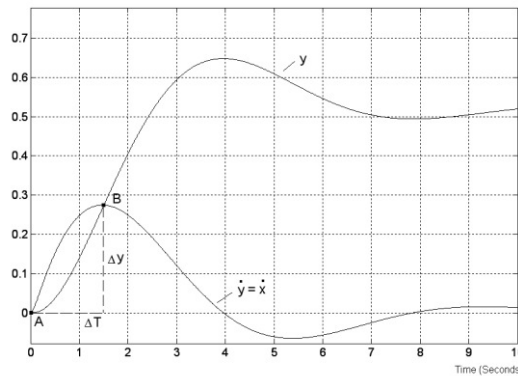


Fig.4 Determination of substitute system acceleration

Example 2: For the function f from Example 1 let us find a function g such that it enables dx to be directly determined by input u , i.e. $dx = u$. The function f is a fuzzy system defined by surface according to Fig. 2. The function g is in fact the function inverse to f with inputs and outputs according to Fig. 5.

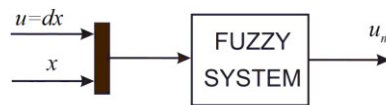


Fig.5 Structure of function $g(u, x)$

When constructing an inverse model of function f , there sometimes occurs the problem of ambiguity [5] in determining its points. This happens in cases when for a single pair $[x, dx]$ there exists several values for u_m as is shown in Fig.6. The solution can define workspaces for inputs of function g that do not contain ambiguity points [5]–[7]. The value u will be in the range $\langle -1, 1 \rangle$ and the value x in the range $\langle 0, 3 \rangle$.

The known relation for function g then serves as a basis for setting up a point matrix to be used for generating its fuzzy description by means of Anfisedit tool of MATLAB. The method used for the fuzzy system generation was the subclustering method with the following parameters:

- Range of influence = 0,5*
- Squash factor = 1,25*
- Accept ratio = 0,5*
- Reject ratio = 0,15.*

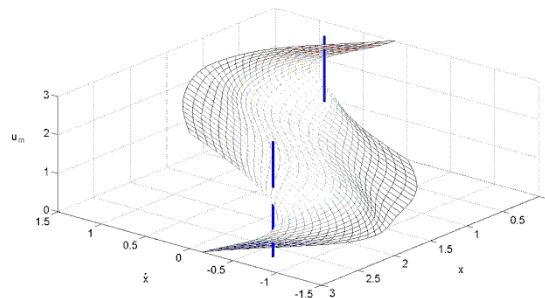


Fig.6 The inverse surface of the system f

Function g is represented by Sugeno-type fuzzy system with 9 rules, the characteristic surface of which is shown in Fig. 7.

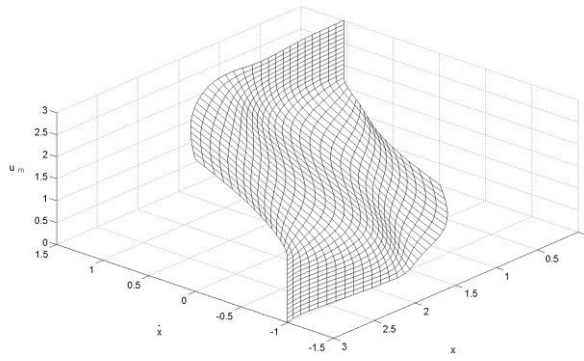


Fig.7 Fuzzy system surface corresponding to function g

V. CONTROL BASED ON AN INVERSE FUZZY MODEL

The basic idea of the control structure based on inverse fuzzy model is shown in Fig. 8. Nonlinearity of the fuzzy system $f(u_m, x)$ is compensated by the nonlinear fuzzy controller $g(dx_{ref}, x)$.

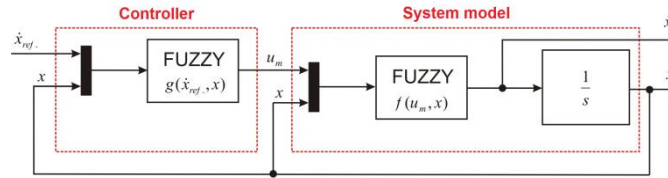


Fig.8 The control structure with inverse fuzzy model

Properties of that control structure with inverse fuzzy model of nonlinear control system (5) have been verified for various required values of variable dx_{ref} by simulation in program Matlab (Fig. 9). At time interval 2-4 seconds we can see a small deviation between required and real value of state variable derivation due to certain inaccuracies in system identification and also inaccuracies in the creation of inverse fuzzy model of the given system. Precise description of the fuzzy function $f(u_m, x)$ as well as the function $g(dx_{ref}, x)$ can eliminate this shortcomings.

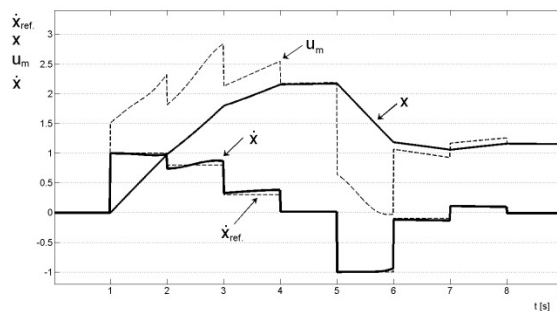


Fig.9 Properties of control structure with inverse fuzzy model

VI. EXPERIMENTAL RESULTS

The method described was used for obtaining a fuzzy model of a drive with asynchronous motor with parameters as per Appendix. The asynchronous motor (AM) represents a fifth-order nonlinear system [8, 9], the structure of which is shown in Fig. 10 where u_x, u_y are components of stator voltage vector U_1 in a rotating coordinate system x, y , and $\omega_1 = 2\pi f_1$ is its angular frequency.

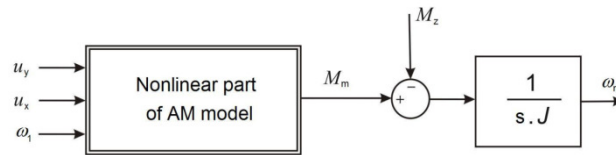


Fig.10 Structure of system with asynchronous motor

The structure of the fuzzy model of a drive with asynchronous motor is illustrated in Fig. 11.

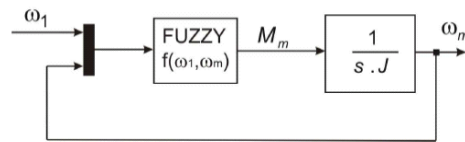


Fig.11 Fuzzy model structure of drive with asynchronous motor

Function f is represented by Sugeno-type fuzzy system with 7 rules, the characteristic surface of which is shown in Fig. 12.

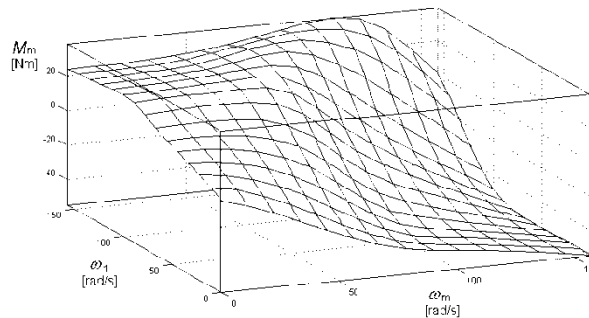


Fig.12 Fuzzy system surface corresponding to system with asynchronous motor

This model was used for the control of the desired torque of an asynchronous motor (M_{m-ref}) in the control structure based on an inverse fuzzy model, as shown in Fig. 13.

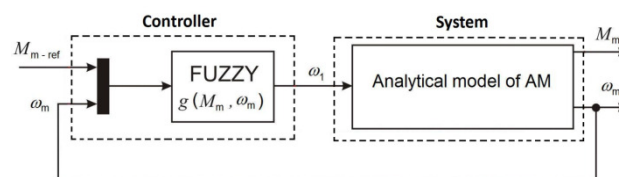


Fig.13 Control structure based on inverse fuzzy model

Function g is the function inverse to function f . When creating an inverse function, it is necessary to define correctly the work space for which all the points of the inverse function g are uniquely determined [5]–[7].

Typical measured control responses of torque and the asynchronous motor angular velocity corresponding to the desired torque are illustrated in Fig. 14 and Fig. 15. The smaller dynamics of the real torque course to the first step in its reference value is caused by the non-excited status of the motor. The resulting quality of regulation strongly depends upon the quality of the fuzzy model of the system, where, due to the application of the described procedure, the fifth-order system becomes a first-order fuzzy system, which in AM causes neglecting of fast electromagnetic activities.

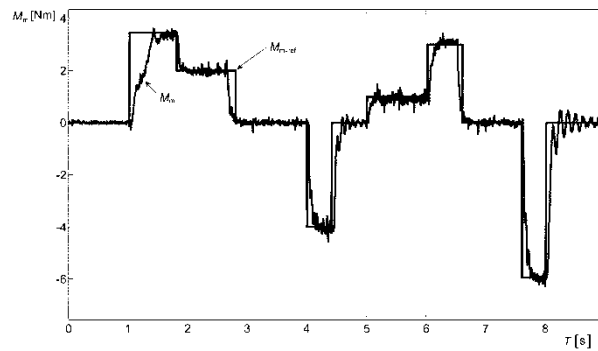


Fig.14 Comparison of measured torque responses with desired values

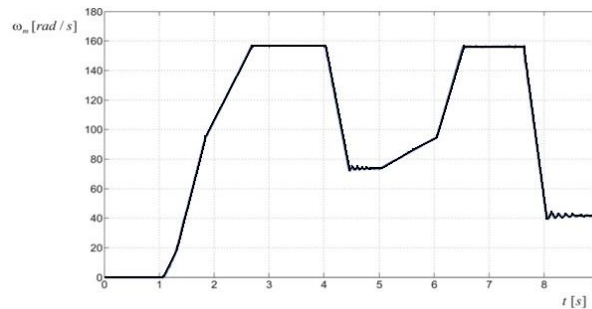


Fig.15 The asynchronous motor angular velocity corresponding to the desired torque

VII. CONCLUSION

The paper presents a method for obtaining a fuzzy model of a nonlinear system from measured data, without necessity of knowledge about its internal structure and parameters. It is demonstrated that this procedure can be fully algorithmized and the fuzzy model obtained can be exploited also for relatively good control of complex nonlinear systems, such as e.g. that of a drive with asynchronous motor. In the future the method described could be applied also to other types of nonlinear systems with similar properties, which render their identification and control using "classical" methods difficult, if not even impossible.

ACKNOWLEDGMENTS

This work was supported by the Slovak Research and Development Agency under the Contract no. APVV-19-0210 and the Contract no. APVV-16-0206.

REFERENCES

- [1] R. Babuška, "Fuzzy Modeling for Control", Kluwer Academic Publishers, Boston, 1998.
- [2] M. Setnes, R. Babuška, H.B. Verbruggen, "Complexity reduction in fuzzy modeling", *IMACS Journal of Mathematics and Computers in Simulation*, vol. 46, 1998, pp. 507-516.
- [3] L. Yan and Y. M. Ma, "Modeling fuzzy information in fuzzy extended entity-relationship model and fuzzy relational databases," *J. of Intelligent & Fuzzy Systems*, vol. 27, no. 4, 2014, pp. 1881-1896. DOI: 10.3233/IFS-141154.
- [4] H. H. Vo, M. Kuchar, and P. Brandstetter, "Application of fuzzy logic in sensorless induction motor drive with PWM-DTC," *Electrical Engineering*, vol. 102, no. 1, 2020, pp. 129-140. DOI: 10.1007/s00202-019-00810-z.
- [5] J. Abonyi, "Fuzzy model identification for control", Birkhäuser, Boston 2003.
- [6] J. Abonyi, H. Andersen, L. Nagy, F. Szeifert, "Inverse Fuzzy-Process-Model Based Direct Adaptive Control", *Mathematics and Computers in Simulation*, vol. 51, Issue 1-2, 1999, pp. 119 – 132.
- [7] C. M. Salgado, J. L. Viegas, C. S. Azevedo, M. C. Ferreira, S. M. Vieira, and J. M. C. Sousa, "Takagi-Sugeno fuzzy modeling using mixed fuzzy clustering," *IEEE Trans. on fuzzy systems*, vol. 25, no. 6, 2017, pp. 1417-1429.
- [8] R.A. Ramadan, "Clustering based fuzzy logic for multimodal sensor networks: A preprocessing to decision fusion," *J. of Ambient Intelligence and Smart Environments*, vol. 3, no. 2, 2010, pp. 271-286. DOI: 10.3233/AIS-2010-0072.
- [9] M. Kuchar, P. Palacky, P. Simonik, and J. Strossa, "Self-Tuning Observer for Sensor Fault-Tolerant Control of Induction Motor Drive," *Energies*, vol. 14, no. 9, 2021, article no. 2564. DOI: 10.3390/en14092564.

Author's Index

A

Anderková Viera 81, 171, 182, 457
Andraško Erik 97
Andrejčík Samuel 62, 92, 128, 159,
224, 307, 325, 409, 485, 523

B

Babič František 81, 171, 182, 457
Bačíková Silvia 347
Baláž Anton 31, 86, 212, 287, 381,
449, 532
Balhová Tímea 437
Bank-Bedei Ivan 200
Baran Gabriel 42
Baumgartner Maroš 134, 218, 330,
476, 496
Bilecová Bianka 335
Binar Martin 369
Bucko Samuel 375
Bučko Jozef 228

C/Č

Culbová Dominika 457
Cymbalak Dávid 241
Cymbalak Dávid 37
Čík Ivan 276

D

Dankanych Natalia 528
Drutarovský Miloš 254, 391
Dudrik Jaroslav 270
Dulinová Júlia 381
Dzúr Dávid 25

F

Fedor Marek 468
Fedor Pavol 552
Fedorco Róbert 234
Figura Radoslav 31
Fortes Jana 443
Frankovič Jakub 311

G

Galovič Ján 295
Gamrát Viktor 512
Gans Šimon 107
Girovský Peter 8, 13, 118, 507, 512
Gladišová Iveta 102
Glevaňák Marek 340
Gojdičová Lucia 113
Golovkin Makar 153
Guzan Milan 301

H

Hasin Martin 295, 340
Havrilla Martin 295, 340
Horčík Tomáš 397
Horváthová Michaela 543
Hudák Marián 54
Hudák Tomáš 415
Hudíková Daniela 194
Hurtuk Ján 146, 335

Ch

Chovancová Eva 74
Chovanec Martin 74, 340

I/J

Ivanchyk Stanislav 165
Jackaninová Miriama 146
Jacko Vladimír 532
Jadlovská Anna 437
Jadlovský Ján 206
Jakab František 37, 241, 311, 352
Jasenka Lukáš 67

K

Kainz Ondrej 311, 352
Kalazi Vojtech 212
Kaňuch Ján 260
Kaňuch Martin 387
Kirešová Simona 301
Kmecik Tadeáš 507
Kolárik Michal 113
Kolcun Michal 123
Korečko Štefan 54, 67, 234
Kormanský Matej 517
Kováč Matej 54
Kucan Michal 37
Kurkina Natalia 134, 218, 476, 496

L

Lacko Dávid 282
Lahunov Dmytro 276
Lapčák Maroš 62, 92, 128, 159, 224,
307, 325, 409, 485, 523
Lonščák Richard 437
Lutak Kristian 421

M

Madoš Branislav 31, 86, 212, 287,
449, 517, 532
Mach Marián 176, 200, 276, 282, 415,
491
Machová Kristína 387, 433, 528, 543,
547
Majirský Daniel 287
Marcinek Adrián 187
Martinko Dávid 123
Matej Tomáš 8
Mattoš Martin 81
Mattová Miriama 54, 67, 234
Medveď Dušan 42, 97, 369
Mičko Kristián 17
Michal Richard 270
Michalko Miroslav 311, 352
Michňák Michal 171
Miščík Tomáš 176
Molnár Ján 107
Mročová Eva 86
Murcko Martin 241
Murín Jaroslav 118

N

Neznik Richard 496

Nováková Ivana 311, 352
Novotná Michaela 249

O

Olšinár Dávid 362
Ondáš Stanislav 249, 347, 427
Onufráková Natália 182
Oravec Jakub 62, 92, 128, 159, 224,
307, 325, 409, 485, 523
Ovseník Luboš 62, 92, 128, 159, 224,
307, 325, 409, 485, 523

P

Papaj Ján 134, 218, 330, 397, 476,
496, 539
Papcun Peter 17
Pavlík Marek 25, 362
Pástor Marek 153, 187, 270, 480
Pekala Patrik 13
Perduková Daniela 468
Petro Viktor 21
Pietriková Emília 381
Pleva Matúš 228
Porezaný Michal 433
Pugelová Zuzana 206

R

Račko Adam 443
Radváni Peter 552
Rohač Marek 391
Rusnák Lubomír 491

S/Š

Sarsala Erik 330
Savinecká Ivana 260
Sedliak Martin 142
Seidenglanz Erik 539
Schwartz Patrik 352
Sobota Branislav 54, 67, 234
Sujová Ema 74
Sulír Matúš 194
Szabó Csaba 403
Szabóová Martina 142, 165
Šimčíková Lea 427
Šimoňák Slavomír 421
Široký Samuel 547

T

Tkáčik Dominik 403
Tóthová Anita 449

V

Vandlík Róbert 403
Vangor Branislav 375
Vaščák Ján 48, 138, 319, 463, 501
Vašková Gabriela 476
Vápeník Roman 37, 241

Z/Ž

Zdravecký Norbert 62, 92, 128, 159,
224, 307, 325, 409, 485, 523
Zeleňák Patrik 254
Žilková Jaroslava 480

Electrical Engineering and Informatics 13

**Proceedings of the
Faculty of Electrical Engineering and Informatics
of the Technical University of Košice**

Published by: Faculty of Electrical Engineering and Informatics
Technical University of Košice
Edition I, 561 pages
Number of CD Proceedings: 50 pieces

Editors: Alena Pietriková
Emília Pietriková

ISBN 978-80-553-4120-0